

BIJIT

Impact Factor 0.605 and Index Copernicus Value (ICV) 4.75

BVICAM'S

International Journal of Information Technology

Indexed with INSPEC (UK), Index Copernicus (Poland), ProQuest (UK), EBSCO (USA), Google Scholar (USA)

CONTENTS

- | | | |
|-----|---|-----|
| 1. | Environmental Intelligence and Its Impact on Collaborative Business
<i>Bharti Trivedi and Bhuvan Unhelkar</i> | 577 |
| 2. | Proposed Fuzzy CPU Scheduling Algorithm (PFCS) For Real Time Operating Systems
<i>Prerna Ajmani and Manoj Sethi</i> | 583 |
| 3. | Web Based Technologies to Support High Level Process Maturity
<i>A. V. Sharmila</i> | 589 |
| 4. | Energy Payback Time Calculation for a Building Integrated Semitransparent Thermal (BISPVT) System with Air Duct
<i>Kanchan Mudgil and Deepali Kamthania</i> | 593 |
| 5. | Effects of Using Filter Based Feature Selection on the Performance of Machine Learners Using Different Datasets
<i>Mehnaz Khan and S. M. K. Quadri</i> | 597 |
| 6. | Path Optimization Using APSO
<i>Deepak Goyal and Malay Ranjan Tripathi</i> | 604 |
| 7. | A Study of Empowerment and Engagement of ITES/BPO Professionals in Delhi-NCR
<i>Puja Khatri and Khushboo</i> | 610 |
| 8. | Cloud Computing and Knowledge Management as a Service: A Collaborative Approach to Harness and Manage the Plethora of Knowledge
<i>Meenu Dave, Mikku Dave and Y. S. Shishodia</i> | 619 |
| 9. | Additional Fault Detection Test Case Prioritization
<i>Ritika Jain and Neha Agarwal</i> | 623 |
| 10. | Short-Term Spectral Feature Extraction and Their Fusion in Text Independent Speaker Recognition: A Review
<i>Ruchi Chaudhary</i> | 630 |
| 11. | Innovative Use of Cloud Computing in Smart Phone Technology
<i>Vaibhav Rana</i> | 640 |
| 12. | A Novel Hénon Map Based Adaptive PSO for Wavelet Shrinkage Image Denoising
<i>Shruti Gandhi, Sonal Doomra, Akash Tayal and Ashwini Kumar</i> | 649 |



Bharati Vidyapeeth's
Institute of Computer Applications and Management
 A-4, Paschim Vihar, Rohtak Road, New Delhi-63

Email : bijit@bvicam.ac.in, Website : <http://www.bvicam.ac.in>

BIJIT - BVICAM's International Journal of Information Technology is a half yearly publication of Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), A-4, Paschim Vihar, Rohtak Road, New Delhi – 110063 (INDIA).

Editor-in-Chief

Prof. M. N. Hoda

Director, BVICAM, New Delhi (INDIA)
E-Mail: bijit@bvicam.ac.in

Editor

Dr. Anurag Mishra

DDU College, University of Delhi
Delhi (INDIA)

Associate Editors

Soft Computing

Prof. K. S. Ravichandran,

Professor and Associate Dean, Dept. of Information and Communication Technologies, Sastra University, Thanjavur – 613401, Tamil Nadu (INDIA)

AI and Innovative Learning Technologies

Dr. Mohamed Hamada,

Senior Associate Professor, Dept. of Computer Science, The University of Aizu, Aizu (JAPAN)

Data Mining, Analytics and Big Data

Dr. Girija Chetty,

Associate Professor, Faculty of Information Technology and Engineering, University of Canberra (AUSTRALIA)

Image Processing

Dr. Pradeep K. Atrey,

Associate Professor, Dept. of Applied Computer Science, The University of Winnipeg (CANADA)

Information Security, High Speed Networks and Cloud Computing

Dr. D. K. Lobiyal,

Associate Professor, School of Computer and Information Sciences, Jawaharlal Nehru University (JNU), New Delhi (INDIA)

Resident Editorial Team

Dr. Deepali Kamthania

Associate Professor, BVICAM
New Delhi (INDIA)

Mr. Vishal Jain

Asstt. Professor, BVICAM
New Delhi (INDIA)

Ms. Ritika Wason

Asstt. Professor, BVICAM
New Delhi (INDIA)

Copy Right © BIJIT – 2013 Vol. 5 No. 2

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical including photocopying, recording or by any information storage and retrieval system, without the prior written permission from the copyright owner. However, permission is not required to copy abstracts of papers on condition that a full reference to the source is given.

ISSN 0973 – 5658

Disclaimer

The opinions expressed and figures provided in the Journal; BIJIT, are the sole responsibility of the authors. The publisher and the editors bear no responsibility in this regard. Any and all such liabilities are disclaimed

All disputes are subject to Delhi jurisdiction only.

Address for Correspondence:

Prof. M. N. Hoda

Editor-in-Chief, BIJIT

Director, Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), A-4, Paschim Vihar, Rohtak Road, New Delhi – 110063 (INDIA). Tel.: 91 – 11 – 25275055
Fax: 91 – 11 – 25255056; E-Mail: bijit@bvicam.ac.in, Visit us at www.bvicam.ac.in/bijit

Published and printed by Prof. M. N. Hoda, Editor-in-Chief, BIJIT and Director, Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), A-4, Paschim Vihar, New Delhi – 110063 (INDIA). Tel.: 91 – 11 – 25275055, Fax: 91 – 11 – 25255056.

E-Mail: bijit@bvicam.ac.in; mca@bvicam.ac.in Visit us at www.bvicam.ac.in/bijit

Our Indexing at International Level



The **INSPEC, IET (UK)**, formerly IEE (UK), database is an invaluable information resource for all scientists and engineers, that contains 13 million abstracts and specialized indexing to the world's quality research literature in the fields of physics and engineering. For further details, click at <http://www.theiet.org/resources/inspec/>



Index Copernicus International (Poland) is a journal indexing, ranking and abstracting site. This service helps a journal to grow from a local level to a global one as well as providing complete web-based solution for small editorial teams. ICV 2012 for the BIJIT is 4.75. For further details, click at

<http://jml2012.indexcopernicus.com/BVICAMs+International+Journal+of+Information+Technology,p4852,3.html>



ProQuest (UK) connects people with vetted, reliable information. Key to serious research, the company has forged a 70-year reputation as a gateway to the world's knowledge – from dissertations to governmental and cultural archives to news, in all its forms. For further details, click at <http://www.proquest.co.uk/en-UK/default.shtml>



EBSCOhost Electronic Journals Service (EJS) is a gateway to thousands of e-journals containing millions of articles from hundreds of different publishers, all at one web site. For further details, click at <http://www.ebscohost.com/titleLists/tnh-coverage.htm>



Open J-Gate is an electronic gateway to global journal literature in open access domain. Launched in 2006, Open J-Gate is aimed to promote OAI. For further details, click at <http://informindia.co.in/education/J-Gate-Engineering/JET-List.pdf>



DOAJ aims at increasing the visibility and ease of use of open access scientific and scholarly journals, thereby promoting their increased usage and impact. For further details, click at

<http://www.doaj.org/doaj?func=issues&jld=87529&uiLanguage=en>



Google Scholar provides a simple way to broadly search for scholarly literature and repositories from across different parts of the world. For further details, click at <http://scholar.google.com/scholar?hl=en&q=BIJIT%2BBVICAM&btnG=>



Cabell's Directory of Publishing Opportunities contains a wealth of information designed to help researchers and academics, match their manuscripts with the scholarly journals which are most likely to publish those manuscripts. For further details, click at <https://ssl.cabells.com/index.aspx>



Academic Journals Database is a universal index of periodical literature covering basic research from all fields of knowledge. For further details, click at <http://journaldatabase.org/journal/issn0973-5658>



Indian Citation Index (ICI) is an abstracts and citation database, with multidisciplinary objective information/knowledge contents from about 1000 top Indian scholarly journals For further details, click at

http://www.indiancitationindex.com/htms/release_notes.htm

and many more..., for more details click at <http://www.bvicam.ac.in/BIJIT/indexing.asp>

Editorial Board

Prof. A. K. Saini

University School of Management Studies, Guru Gobind Singh Indraprastha University, New Delhi (INDIA)

Prof. A. K. Verma

Centre for Reliability Engineering, IIT Mumbai, Mumbai (INDIA)

Prof. A. Q. Ansari

Dept. of Electrical Engineering, Jamia Millia Islamia, New Delhi (INDIA)

Dr. Amudha Poobalan

Division of Applied Health Sciences, University of Aberdeen, Aberdeen (UK)

Prof. Anand Bhalerao

Dept. of Civil Engineering, Bharati Vidyapeeth's College of Engineering, Pune (INDIA)

Prof. Anwar M. Mirza

Dept. of Computer Science, National University of Computer & Emerging Sciences, Islamabad (PAKISTAN)

Prof. Ashok K. Agrawala

Dept. of Computer Science, Director, The MIND Lab and The MAXWell Lab, University of Maryland, Maryland (USA)

Prof. B. S. Chowdhry

Dept. of Electronics Engineering, Mehran University of Engineering & Technology (PAKISTAN)

Dr. Bimlesh Wadhwa

School of Computing, National University of Singapore, Singapore (JAPAN)

Prof. Clarence Wilfred DeSilva

Dept. of Mechanical Engineering, University of British Columbia (CANADA)

Prof. D. K. Bandyopadhyay

Vice Chancellor, Guru Gobind Singh Indraprastha University, New Delhi (INDIA)

Dr. D. M. Akbar Hussain

Dept. of Energy Technology, Aalborg University, Esbjerg (DENMARK)

Prof. David L Olson

Dept. of Management, University of Nebraska (USA)

Dr. Fahim Mohammad

Harvard Medical School, Harvard University, Boston (USA)

Prof. Hamid R. Arabnia

Dept. of Computer Science, University of Georgia (USA)

Dr. Hasmukh Morarji

School of Software Engineering and Data Communications, Queensland University of Technology, Brisbane (AUSTRALIA)

Dr. Javier Poncela

Dept. of Electronic Technology, University of Malaga (SPAIN)

Prof. K. K. Aggarwal

Former Vice Chancellor, Guru Gobind Singh Indraprastha University, New Delhi (INDIA)

Prof. K. Poulouse Jacob

Dept. of Computer Science, University of Science and Technology, Cochin (INDIA)

Prof. Ken Surendran

Dept. of Computer Science, Southeast Missouri State University, Cape Girardeau Missouri (USA)

Dr. Ki Young Song

Dept. of Mechanical Engineering, The University of Tokyo, Tokyo (JAPAN)

Prof. Kishor Trivedi

Dept. of Electrical and Computer Engineering , Duke University (USA)

Prof. M. N. Doja

Dept. of Computer Engineering, Jamia Millia Islamia, New Delhi (INDIA)

Prof. M. P. Gupta

Dept. of Management Studies, IIT Delhi, New Delhi (INDIA)

Prof. Madan Gupta

Director, Intelligent Systems Research Laboratory, University of Saskatchewan, Saskatoon, Saskatchewan (CANADA)

Prof. Mohammad Yamin

School of Computer Science, The Australian National University, Canberra (AUSTRALIA)

Dr. Nurul Fadly Bin Habidin

Engineering Business and Management, University Pendidikan Sultan Idris (MALAYSIA)

Prof. O. P. Vyas

Dept. of Information Technology, Indian Institute of Information Technology Allahabad (IIITA), Allahabad (INDIA)

Prof. Prasant Mohapatra

Dept. of Computer Science, University of California (USA)

Prof. Richard Chbeir

School of Computer Science, Université de Pau et des Pays de l'Adour (UPPA), Anglet (FRANCE)

Dr. S. Arockiasamy

Dept. of Information Systems, University of Nizwa, Sultanate of Oman (OMAN)

Prof. S. I. Ahson

Former Pro-Vice-Chancellor, Patna University, Patna (INDIA)

Prof. S. K. Gupta

Dept. of Computer Science and Engineering, IIT Delhi, New Delhi (INDIA)

Prof. Salim Beg

Dept. of Electronics Engineering, Aligarh Muslim University, Aligarh (INDIA)

Prof. Shiban K. Koul

Centre for Applied Research in Electronics (CARE), IIT Delhi, New Delhi (INDIA)

Prof. Shuja Ahmad Abbasi

Dept. of Electrical Engineering, King Saud University, Riyadh (KSA)

Prof. Steven Guan

Dept. of Computer Science & Software Engineering, Xi'an Jiaotong-Liverpool University (CHINA)

Prof. Subir Kumar Saha

Dept. of Mechanical Engineering, IIT Delhi, New Delhi (INDIA)

Prof. Subramaniam Ganesan

Dept. of Computer Science and Engineering, Oakland University, Rochester (USA)

Prof. Susantha Herath

School of Electrical and Computer Engineering, St. Cloud State University, Minnesota (USA)

Prof. Yogesh Singh

Vice Chancellor, MS University, Baroda (INDIA)

Editorial

It is a matter of both honor and pleasure for us to put forth the tenth issue of BIJIT; the BVICAM's International Journal of Information Technology. It presents a compilation of twelve papers that span a broad variety of research topics in various emerging areas of Information Technology and Computer Science. Some application oriented papers, having novelty in application, have also been included in this issue, hoping that usage of these would further enrich the knowledge base and facilitate the overall economic growth. This issue again shows our commitment in realizing our vision "to achieve a standard comparable to the best in the field and finally become a symbol of quality".

As a matter of policy of the Journal, all the manuscripts received and considered for the Journal, by the editorial board, are double blind peer reviewed independently by at-least two referees. Our panel of expert referees posses a sound academic background and have a rich publication record in various prestigious journals representing Universities, Research Laboratories and other institutions of repute, which, we intend to further augment from time to time. Finalizing the constitution of the panel of referees, for double blind peer review(s) of the considered manuscripts, was a painstaking process, but it helped us to ensure that the best of the considered manuscripts are showcased and that too after undergoing multiple cycles of review, as required.

The twelve papers that were finally published were chosen out of ninety eight papers that we received from all over the world for this issue. We understand that the confirmation of final acceptance, to the authors / contributors, sometime is delayed, but we also hope that you concur with us in the fact that quality review is a time taking process and is further delayed if the reviewers are senior researchers in their respective fields and hence, are hard pressed for time.

We further take pride in informing our authors, contributors, subscribers and reviewers that the journal has been indexed with some of the world's leading

*indexing / bibliographic agencies like INSPEC of IET (UK) formerly IEE (UK), Index Copernicus International (Poland) with IC Value 4.75 for 2012, ProQuest (UK), EBSCO (USA), Open J-Gate (USA), DOAJ (Sweden), Google Scholar, WorldCat (USA), Cabell's Directory of Computer Science and Business Information System (USA), Academic Journals Database, Open Science Directory, Indian Citation Index, etc. and listed in the libraries of the world's leading Universities like Stanford University, Florida Institute of Technology, University of South Australia, University of Zurich, etc. Related links are available at <http://www.bvicam.ac.in/bijit/indexing.asp>. Based upon the papers published in the year 2012, its **Impact Factor** was found to be **0.605**. These encouraging results will certainly further increase the citations of the papers published in this journal thereby enhancing the overall research impact.*

We wish to express our sincere gratitude to our panel of experts in steering the considered manuscripts through multiple cycles of review and bringing out the best from the contributing authors. We thank our esteemed authors for having shown confidence in BIJIT and considering it a platform to showcase and share their original research work. We would also wish to thank the authors whose papers were not published in this issue of the Journal, probably because of the minor shortcomings. However, we would like to encourage them to actively contribute for the forthcoming issues.

The undertaken Quality Assurance Process involved a series of well defined activities that, we hope, went a long way in ensuring the quality of the publication. Still, there is always a scope for improvement, and so, we request the contributors and readers to kindly mail us their criticism, suggestions and feedback at bijit@bvicam.ac.in and help us in further enhancing the quality of forthcoming issues.

Editors

CONTENTS

1.	Environmental Intelligence and Its Impact on Collaborative Business	577
	<i>Bharti Trivedi and Bhuvan Unhelkar</i>	
2.	Proposed Fuzzy CPU Scheduling Algorithm (PFCS) For Real Time Operating Systems	583
	<i>Prerna Ajmani and Manoj Sethi</i>	
3.	Web Based Technologies to Support High Level Process Maturity	589
	<i>A. V. Sharmila</i>	
4.	Energy Payback Time Calculation for a Building Integrated Semitransparent Thermal (BISPVT) System with Air Duct	593
	<i>Kanchan Mudgil and Deepali Kamthania</i>	
5.	Effects of Using Filter Based Feature Selection on the Performance of Machine Learners Using Different Datasets	597
	<i>Mehnaz Khan and S. M. K. Quadri</i>	
6.	Path Optimization Using APSO	604
	<i>Deepak Goyal and Malay Ranjan Tripathi</i>	
7.	A Study of Empowerment and Engagement of ITES/BPO Professionals in Delhi-NCR	610
	<i>Puja Khatri and Khushboo</i>	
8.	Cloud Computing and Knowledge Management as a Service: A Collaborative Approach to Harness and Manage the Plethora of Knowledge	619
	<i>Meenu Dave, Mikku Dave and Y. S. Shishodia</i>	
9.	Additional Fault Detection Test Case Prioritization	623
	<i>Ritika Jain and Neha Agarwal</i>	
10.	Short-Term Spectral Feature Extraction and Their Fusion in Text Independent Speaker Recognition: A Review	630
	<i>Ruchi Chaudhary</i>	
11.	Innovative Use of Cloud Computing in Smart Phone Technology	640
	<i>Vaibhav Rana</i>	
12.	A Novel Hénon Map Based Adaptive PSO for Wavelet Shrinkage Image Denoising	649
	<i>Shruti Gandhi, Sonal Doomra, Akash Tayal and Ashwini Kumar</i>	

Environmental Intelligence and its Impact on Collaborative Business

Bharti Trivedi¹ and Bhuvan Unhelkar²

Submitted in March, 2013; Accepted in August, 2013

Abstract - Green IT can be defined as the conscious implementation of technologies, techniques and policies designed to reduce the carbon footprint of the IT function within the organization and the use of IT to reduce the carbon footprint of the whole organization. By adopting Green IT practices, business enterprises can contribute positively to environmental stewardship and protect the environment while reducing the energy costs. The holistic approach to environmental sustainability making the best use of IT resources available to the organization is termed as environmental intelligence (EI) in this paper.

This paper explores creative ways of looking at the environmental challenge, opens up opportunities to examine processes for collaboration, take stock of the inventory and infrastructure for optimization, and explore the possibilities of new business streams. This research paper aims to find out the impact of technologies such as server virtualization, cloud computing, anytime anywhere technologies and right network (as a specialist domain of Green ICT) on EI of different business organizations of Vadodara. This study employed the ordinal logistic regression to understand the impact of green ICT on the manufacturing business.

Index Terms - Green IT, Environmentally Sustainable, Environmental Intelligence (EI), Server Virtualization, Cloud Computing

1. INTRODUCTION

A sensible and carefully created environmental strategy will not only handle the immediate environmental impact but will also include carbon performance as a part of its risk management approach. The environmentally conscious practices of an organization are not restricted to its IT equipments and process; instead this becomes an all-encompassing adventure by its organization that also includes the consideration for emerging ICT in the business processes and policies [1]. Green IT is studied from two angles: reducing the emissions from IT use in business and using IT to reduce the carbon footprints of an organization. The primary value of Green IT is derived through environmental performance improvement in all sectors of an organization by using IT products and technologies.

¹ DDU, Nadiad, India,

² Consultant, MethodScience.com, University of Western Sydney, Australia

E-mail: ¹bhartiapex@yahoo.co.in and

²bhuvan@methodscience.com

The Green IT thus also creates profitable opportunities for business by enabling design and deployment of IT-based solutions that create a low carbon economy [2]. The IT industry has, of late, extended the scope of Green IT to include the manner in which IT is deployed to help reduce GHG emissions, to introduce energy efficiencies, or to reduce energy consumption [3] [4].

This research paper studies the impact of environmental Intelligence (EI) for Green ICT on the collaborative business ecosystem of Vadodara. Vadodara is the third most populated city in the Indian state of Gujarat. In line with the 'Knowledge City' vision of the confederation of Indian Industry, Vadodara is gradually becoming a hub in Gujarat for IT and other development projects. A survey was conducted using cross sectional approach. The research population consist of all the business units of Vadodara which are hugely using IT and IT infrastructure in the organization.

This research has explored the current application of Green ICT as well as the readiness for Green ICT in an organization. The research gave a new insight to EI, by explaining EI as a practical approach bringing together the tools and techniques of Green ICT to achieve the dual purpose of business and environmental efficiency. This research has investigated the impact of EI on the different business categories.

2. LITERATURE REVIEW

Green ICT is concerned with both the viability of the organization and its efficient operations [5]. Osborne [6] stated that attainment of Green IT is that optimum balance between minimizing the responsibility for carbon emissions while delivering the optimum IT function to support the enterprise. This cannot be zero; the equipment life cycle will generate an impact, but steps can be taken to minimize the impacts. The IT sector has a profitable opportunity and a critical role to play with other sectors to design and deploy solutions needed to create a low carbon society [7]. This research explores the depth of environmental challenges in the current business setup through emerging ICT. Right network, Anytime anywhere technologies, server virtualization and cloud computing are the technologies which has the ability to reduce an organization's carbon footprint by facilitating more efficient and less carbon intensive work practices such as teleconferencing instead of flying or commuting [8].

Green IT is expected to improve both environmental and economic challenges. The emerging technologies of IT focus on the opportunities and have the ability to transform the current business model to EI business. Cloud computing has emerged as the natural evolution and integration of advances in several fields including utility computing, distributed

computing, grid computing, web services, and service oriented architecture [9]; providing substantial opportunities for organizations to consolidate their hardware and corresponding data space requirements [10] . This is a whole new form of computing and is allowing thousands of users from all around the world to access something without having to download and install anything on their own computers.

Mobility enables virtual collaboration between business and individuals. Reengineering the business processes with mobility provides enormous opportunities for virtualization. The more virtual a business is the less physical resources it will consume—therefore, well-modeled mobile processes greatly assist in creation of environment friendly businesses. Mobility has the potential to further enhance the data sharing among the business application. This is so because mobility enables location independent data sharing. Mobility reduces access time, optimizes input of data, and offers location-based insights. The mobile devices have the capability to connect to the global network while they are connected to individual people, regardless of their location and time. Mobility is a significant factor in the quality of life of individuals and society as a whole [11].

The ICT technologies discussed above can be applied to the following business purposes, in order to drive environmental intelligence (EI); Reduction of energy consumption, reduction of carbon footprints, reduction of the operational costs are the important goals defined by organizations to adopt green practices and policies for the Green ICT.

3. RESEARCH MODEL AND HYPOTHESIS DEVELOPMENT

Based on previous studies a questionnaire was developed to understand the impact of Green IT on the environmental goals of the business. The first section of the questionnaire consisted of items such as name of the organization, number of computers and peripherals, information about computer networking. The second section included the Green ICT factors: Right Network (RN), Anytime Anywhere Technologies (ATAW), Server Virtualization (SV), and Cloud Computing (CC) influencing the environmental goals of a business. The third section included variables which measure the dependent variable. Impact is identified by reduction of energy consumption, reduction of carbon footprints and reduction of operational cost [12]. Finally all the three variables summarize into a single variable, which is a dependent variable (DV) called 'Impact of Environmental Intelligence' of a business from expert's viewpoint [26]. Independent variables can be categorized into four factors which are detailed along with the hypothesis in Table 1.

The hypothesis in Table 1 are detailed as -

H1: There is a significant impact of use of EI for RN in collaborative business.

H2: There is a significant impact of use of EI for ATAW in collaborative business.

H3: There is a significant impact of use of EI for SV in collaborative business.

H4: There is a significant impact of use of EI for CC in collaborative business.

Factors	Items	Hypothesis
RN	Global network protocol , Network can be easily scalable to support growth , Supports collaborations and mobile technologies , Supports effective supply, production and distribution processes centralization and better performance of line-of-business applications ,Supports effective ecommerce , Handles all voice, data and multimedia traffic and applications on one network eliminating duplication of bandwidth, networking equipment and network management resources	H1
ATAW	Minimize the need to travel at all via remote working and video conferencing to reduce congestion and vehicle emission, Make presentations to customers, and being able to download product information to their network during the visit, Quote business quotations and interactive order processing , Check stock levels via the office network , Interact with colleagues while travelling - sending and receiving emails, collaborating on responses to tenders, delivering trip reports in a timely manner.	H2
SV	Reduce hardware peripherals, Share applications instances between multiple organizations, Improve time-to-market of new applications , Better sharing infrastructure investments across groups, Operational system efficiency and simplified management due to better use of infrastructure ,Be more responsive to the needs of the business ,Improve business continuity and disaster recovery ,Speed up deployment of new applications	H3
CC	Data Duplication , Access control , Innovation , Outsourcing , Tele-working , Efficient data handling, Increase technical capability	H4

Table 1: Independent Variables (IV) and Research Hypothesis

4. METHODOLOGY

The research subject “to investigate the impact of use of EI for Green IT practices on an organization” predetermines the choice of statistical test and analysis to be used in the study.

389 business enterprises constitute the population; the population is divided into 8 strata. If N is the size of the population, n is the size of the sample and i represents 1,2,...k [number of strata in the population], then proportionate stratified sampling is defined as

$$p = n1/N1 = n2/N2 = \dots n8/N8 = n/N,$$

p = 70% is the proportion selected from each strata in the population in this research (as shown in Table 2). The stratum are manufacturing units (S1), Schools (S2), Colleges(S3), Hospitals (S4), Banks, Finance and Insurance (S5), Retail and Shopping Malls(S6), Administrative and Support services (S7) and Computer training and Software development centers (S8).

Stratum No.	Stratum population	Stratum Sample selected (70%)	Sample Response
S1	73	51	50
S2	65	45	40
S3	32	23	20
S4	37	26	22
S5	42	29	21
S6	40	28	21
S7	44	30	19
S8	54	37	32

Table 2: Stratum Population and Stratum Sample

A 5–point likert scale ranging from 1 as strongly disagrees to 5 as strongly agree was used for the measurement. A test for the reliability of the instrument was conducted for each stratum population. The internal consistency is checked using cronbach’s alpha coefficient and is tabulated in Table 3.

Stratum	RN	ATAW	SV	CC
S1	0.8324	0.7255	0.7045	0.6982
S2	0.7711	0.7232	0.8722	0.8223
S3	0.8712	0.7342	0.7611	0.7433
S4	0.8732	0.8732	0.8345	0.7764
S5	0.7834	0.8876	0.8342	0.7943
S6	0.7223	0.8811	0.4672	0.3588
S7	0.7056	0.7232	0.7722	0.7223
S8	0.8722	0.8762	0.7722	0.8743

Table 3: Reliability Test

This research employed the ordinal logistic regression to drive its results. An ordinal logistic regression is used to handle multiple ordinal dependent variables. Minitab 14 is a statistical software package that can fit an ordinal logistic regression to data. The output of the software includes: 1) Response and Factor Information, which displays the number of observations and the response and factor categories; 2) Logistic Regression Table, which shows the estimated coefficients (C), p-values (P) (related to a test that the corresponding coefficient is zero) and

odds ratio (O) (which shows the effect of variables on the model); 3) Goodness-of-Fit Tests which displays both Pearson goodness-of-fit test of the model to data. Final and appropriate model is chosen by entering variables which their coefficients are significant (p-value<0.05) and ordering effect of variables from their odds ratio negative coefficient along smallest odds ratio indicates more impact of the variable on the dependent variable [13]. Finally, appropriative of model is evaluated by (i) a G test whose null hypothesis states all the coefficients associated with predictors equal zero versus at least one coefficient is not zero (we prefer to reject their null hypothesis, i.e., p-value <0.05) and (ii) Goodness-of-Fit Tests, (we prefer to accept their null hypothesis, i.e., p-value >0.05), of which more can be found [13]. The equation of ordinal logistic regression in this research model are given below.

$$\gamma_i = \frac{\exp(\alpha_i + \text{coeff}(RN) + \text{coeff}(ATAW) + \text{coeff}(SV) + \text{coeff}(CC))}{1 + \exp(\alpha_i + \text{coeff}(RN) + \text{coeff}(ATAW) + \text{coeff}(SV) + \text{coeff}(CC))}$$

where i= (1,2,3,4)

γ_i is the cumulative probability efficiency of the EI of ith level ,

α_i are the coefficients for each level.

5. RESULTS

The impact of environmentally responsible business strategies in the businesses are tabulated in Table 4 to Table 7. The odds that an event occurs is the ratio of the number of people who experience the event to the number of people who do not. This is what one get when one divide the probability that the event occurs by the probability that the event does not occur, since both probabilities have the same denominator and it cancels, leaving the number of events divided by the number of non-events. The coefficients in the logistic regression model tell how much the logic changes based on the values of the predictor variables.

5.1 EI and its impact on reduction of energy consumption in business

Table 4 shows the results of the Ordinal Logistic Regression for the impact of EI on reduction of energy consumption in all the strata.

		RN	ATAW	SV	CC
S1	C	-0.351	-0.332	-0.772	-1.67
	P	0.046	0.003	0.016	0.028
	O	0.69	0.28	0.55	0.67
	R	4	1	2	3
S2	C	-0.887	-1.876	-0.781	-0.771
	P	0.001	0.089	0.004	0.453
	O	0.43	1.65	0.87	0.66
	R	1	-	2	-
S3	C	-0.112	-1.665	-2.555	-0.0112
	P	0.022	0.000	0.000	0.016
	O	0.71	0.56	0.62	0.67
	R	4	1	2	3
S4	C	-0.228	-0.778	-1.778	-1.778

		RN	ATAW	SV	CC
	P	0.032	0.214	0.074	0.397
	O	0.24	1.22	0.77	0.76
	R	1	-	-	-
S5	C	-0.213	-1.778	-0.671	-2.562
	P	0.046	0.003	0.009	0.006
	O	0.41	0.28	0.38	0.32
	R	4	1	3	2
S6	C	-0.223	-0.343	-	-
	P	0.018	0.003	-	-
	O	0.66	0.32	-	-
	R	2	1	-	-
S7	C	-0.786	-0.771	-0.771	-2.881
	P	0.036	0.003	0.012	0.028
	O	0.67	0.22	0.39	0.51
	R	4	1	2	3
S8	C	-0.662	-1.661	-0.191	-0.881
	P	0.046	0.003	0.016	0.028
	O	1.92	0.67	1.87	1.89
	R	4	1	2	3

Table 4: EI and its impact on Reduction of Energy Consumption in Business

Results in Table 4 summarizes that p-value (>0.05) indicates that businesses are not adopting such practices and hence they do not contribute to the reduction of energy consumption. A comparative study of collaborative businesses shows that stratum 2, stratum 4 and stratum 6 has least awareness regarding use of emerging ICT & has insignificant impact on the energy consumption of the business. The negative coefficients and the p-values (<0.05) for the choice of right network indicates that right network practices are related to lower the energy consumption of the business.

5.2 EI and its impact on reduction of carbon footprints in business

Table 5 shows the impact on EI on the reduction of carbon footprints of a business organization.

		RN	ATAW	SV	CC
S1	C	0.291	-0.561	-0.721	0.82
	P	0.649	0.002	0.018	0.16
	O	0.67	0.29	1.1	1.7
	R	-	1	2	-
S2	C	0.657	0.897	0.887	0.11
	P	0.339	0.074	0.148	0.16
	O	0.99	0.89	0.98	0.87
	R	-	-	-	-
S3	C	1887	-1.65	-1.9	-0.886
	P	0.546	0.001	0.009	0.028
	O	1.9	0.37	0.66	0.98
	R	-	1	2	3
S4	C	0.912	0.877	4.2	1.98
	P	0.395	0.074	0.397	0.134

		RN	ATAW	SV	CC
	O	1.76	1.98	0.88	0.56
	R	-	-	-	-
S5	C	-0.871	-1.769	0.887	129
	P	0.003	0.002	0.214	0.047
	O	1.65	0.54	1.92	1.71
	R	2	1	-	3
S6	C	0.881	-2.109	-	-
	P	0.649	0.002	-	-
	O	0.76	0.76	-	-
	R	-	1	-	-
S7	C	4.76	1.223	-0.564	0.871
	P	0.623	0.392	0.014	0.187
	O	0.88	0.39	0.38	0.66
	R	-	-	1	-
S8	C	-0.619	-0.672	-0.781	-0.261
	P	0.016	0.002	0.018	0.002
	O	2.1	1.01	2.9	1.23
	R	3	1	4	2

Table 5: EI and its impact on reduction of carbon footprints in business

Results in Table 5 can be summarized as most businesses are not well conversant with carbon reduction and carbon footprints from the use of ICT. Reduction of carbon emissions is not a strong motivational factor in businesses to adopt EI.

5.3 EI and its impact on reduction of operational cost in business

Table 6 comprehensively illustrates the impact of EI on the reduction of operational cost in business.

		RN	ATAW	SV	CC
S1	C	-1.262	0.810	1.887	0.87
	P	0.018	0.075	0.166	0.075
	O	0.72	0.77	0.77	0.77
	R	1	-	-	-
S2	C	0.8978	0.291	0.828	-0.675
	P	0.318	0.075	0.117	0.017
	O	2.1	1.8	0.89	0.38
	R	-	-	-	1
S3	C	-0.651	-0.78	-1.818	-0.672
	P	0.018	0.032	0.005	0.012
	O	1.89	2.05	0.78	1.1
	R	3	4	1	2
S4	C	0.261	0.2611	0.191	0.987
	P	0.649	0.075	0.166	0.214
	O	0.78	0.87	0.27	0.76
	R	-	-	-	-
S5	C	-1.8	-0.272	-1.191	-0.538
	P	0.018	0.003	0.018	0.001
	O	1.1	0.78	1.5	0.61
	R	3	2	4	1
S6	C	-0.87	0.819	-	-
	P	0.006	0.075	-	-

		RN	ATAW	SV	CC
	O	0.69	1.4	-	-
	R	1	-	-	-
S7	C	-2.001	1.982	0.672	1.89
	P	0.018	0.075	0.107	0.075
	O	0.41	0.76	0.78	0.82
	R	1	-	-	-
S8	C	-0.502	-0.366	-0.464	-0.119
	P	0.018	0.002	0.002	0.009
	O	0.74	0.29	0.36	0.51
	R	4	1	2	3

Table 6: EI and its impact on reduction of operational cost in business

The reduction of operational cost is highlighted as one of the important motivation for reducing energy efficient ICT. The results in Table 6 shows that the use of emerging ICT such as ATAW technologies, CC are correlated to the reduction of operational cost such as reduction of paper and travelling.

5.4 EI and its impact on the collaborative business

All the variables are ordinal; a median is employed to summarize a group of items into one single variable. They all grouped in one variable called Impact. Table 7 summarizes EI and its impact on collaborative business.

		RN	ATAW	SV	CC
S1	C	*	-1.54	*	*
	P	0.102	0.006	0.429	0.28
	O	*	0.37	*	*
	R	*	1	*	*
S2	C	-1.743	*	*	-0.629
	P	0.013	0.156	0.071	0.016
	O	0.69	*	*	0.78
	R	1	*	*	2
S3	C	-0.321	-0.719	-2.161	-0.811
	P	0.005	0.003	0.021	0.025
	O	0.38	0.29	1.39	1.42
	R	2	1	3	4
S4	C	*	*	*	*
	P	0.072	0.56	0.167	0.181
	O	*	*	*	*
	R	*	*	*	*
S5	C	-2.814	-2.371	*	-0.562
	P	0.005	0.002	0.721	0.037
	O	0.37	0.29	*	1.15
	R	2	1	*	
S6	C	-0.924	*	*	*
	P	0.032	0.281	0.551	0.072
	O	2.01	*	*	*
	R	1	*	*	*
S7	C	-0.945	*	*	*
	P	0.015	0.62	0.81	0.055
	O	0.82	*	*	*
	R	1	*	*	*
S8	C	-2.12	-0.453	-0.591	-0.671

		RN	ATAW	SV	CC
	P	0.014	0.002	0.016	0.003
	O	1.15	0.67	1.18	0.72
	R	3	1	4	2

Table 7: EI and its impact on the collaborative Business

If the p-value is less than the alpha-level (0.05), the test rejects the null hypothesis that the model does not fit the data adequately and accepts the hypothesis. In Table 8 √ represents that the hypothesis is accepted and × represents that the hypothesis is rejected. The hypothesis is tested stratum wise.

	H1	H2	H3	H4
S1	×	√	×	×
S2	√	×	×	√
S3	√	√	√	√
S4	×	×	×	×
S5	√	√	×	√
S6	√	×	×	×
S7	√	×	×	×
S8	√	√	√	√

Table 8: EI and its impact on Collaborative Business

6. CONCLUSION

The study suggests that Environmental Intelligence for greening of ICT and its application is less explored in Vadodara region, the business sections covering hospitals and retail shopping malls are not considering any significant impact of greening of ICT in their business. Right network and anytime any where technologies have more impact on the businesses. Motivational factors are the reduction of energy consumption and reduction of operational cost.

The results tabulated and analyzed in Section 5.0 depicts that Environmental Intelligence (EI) in computing represents a responsible way to address the issue of energy consumption and operational cost in a business enterprise. By adopting EI practices, business leaders can contribute positively to environmental stewardship and reduce environmental footprints while reducing energy and operational cost.

Based on the results, it can be recommended that strategic alignment of new and emerging ICT technologies such as mobile technology, cloud computing, virtualization with business has been a key in delivering competitive advantage to business, especially in stratum 1(manufacturing units), stratum 3 (Colleges), stratum 5 (Banks, Finance & insurance companies)and stratum 8 (Computer Training & Software Development Centers). This alignment needs to be kept in mind when it comes to innovative use of emerging technologies and carbon reduction.

The technologies that are most likely to have an impact on collaborative business are right network and mobile technologies. The results in Section 5.0 illustrate that cloud computing is now becoming an important part of an organization’s business approach, but they also require a corresponding business model that can support the use of these technologies.

REFERENCES

- [1]. B. Unhelkar, "Green IT Fundamentals", Green IT Strategies and Applications using Environmental Intelligence, CRC Press, Taylor and Francis Group, ISBN: 978-1-4398-3780-1, 2011.
- [2]. B. Trivedi & B. Unhelkar, "Managing Environmental Compliance: A Techno-Business Perspective", SCIT Journal , Chapter 2, pp 10- 17, Volume IX, August 2009
- [3]. B. Adjugo, "Green IT Survey 2007 – 2008", Green IT Governance: Be Part of a Cultural Change, www.Green-ICT.com, 2008
- [4]. Gartner Research ,Gartner estimates ICT industry accounts for two percent of Global Co2 emissions, 2007, Press release, Gartner Newsroom, April 26, 2007, Stamford, Retrieved August 16, 2009, from <http://www.gartner.com/it/page.jsp?Id=503867>
- [5]. M. C. Boudreau, R.T. Watson, and A. Chen, "From Green IT to Green IS". In Biros, B. et al., The Organizational Benefits of Green IT, Cutter Information LLC, Arlington, MA, 79-91, 2008..
- [6]. Osborne, "The Green Data Centre: Taking the first steps towards Green IT", Cutter Consortium, pp 125-134, 2008.
- [7]. Y. Lan. & P.T. Hywel, "A Review of Research on the Environment Impact of E-Business and ICT", Environment International, Vol 33, No 6, August 2007, 2007, pp 841-849
- [8]. B. Trivedi, "Green Enterprise Architecture using Environmental Intelligence", International Journal of Green Computing, Vol2, No 1, Jan-June , 2011, ISSN:1948-5018, 2011
- [9]. Weiss, "Computing in the clouds", ACM netWorker 11(4):16-25, Dec. 2007.
- [10]. S. Murugesan, Cloud Computing, Chapman-Hall, USA, 2011
- [11]. B. Unhelkar, "Transition to a Mobile Enterprise: A Three Dimensional Framework", Cutter IT Journal, special issue on Mobile Computing, ed. S. Murugesan. 18(8), 2005
- [12]. B. Unhelkar, & B. Trivedi, "Merging Web Services With 3G IP Multimedia Systems For Providing Solutions In Managing Environmental Compliance By Business",Proceedings of ITA09, Wrexham, UK, 8th Sep 2009 to 11 Sep 2009
- [13]. P. McCullagh & J.A. Nelder, Generalized Linear Model. Chapman & Hall, 1992.

Proposed Fuzzy CPU Scheduling Algorithm (PFCS) for Real Time Operating Systems

Perna Ajmani¹ and Manoj Sethi²

Submitted in March, 2013; Accepted in August, 2013

Abstract - *In the era of supercomputers multiprogramming operating system has emerged. Multiprogramming operating system allows more than one ready to execute processes to be loaded into memory. CPU scheduling is the process of selecting from among the processes in memory that are ready to execute and allocate the processor time (CPU) to it. Many conventional algorithms have been proposed for scheduling CPU such as FCFS, shortest job first (SJF), priority scheduling etc. But no algorithm is absolutely ideal in terms of increased throughput, decreased waiting time, decreased turnaround time etc. In this paper, a new fuzzy logic based CPU scheduling algorithm has been proposed to overcome the drawbacks of conventional algorithms for efficient utilization of CPU.*

Index Terms - *CPU scheduling, fuzzy logic, Multiprogramming Operating System, process, turnaround time, and throughput.*

1. INTRODUCTION

With the advancement in operating system, multiprogramming operating systems has evolved. In a multiprogramming environment, many processes are loaded into memory that competes for CPU time. CPU scheduling algorithms determines which process will be given processor time and which will wait. Some of the objectives that scheduling function should satisfy in order to be effective include fairness, efficient use of processor time, response time, turnaround and throughput [11]. There are many scheduling algorithms such as FCFS, SJF, PRIORITY Scheduling etc., but none is efficient for real time tasks.

1. **FCFS**: - In FCFS algorithm the process is allotted processor time on First Come, First Serve basis. It is a non-preemptive scheduling in which the processes are being given CPU in the order of their arrival in ready queue. Advantage of FCFS is less context switching overhead. But the limitations are: - (i) Throughput can be low, since long processes can hold the CPU. (ii) Turnaround time, waiting time and response time can be high for the same reason. (iii) No prioritization occurs, thus this system has trouble to meet deadlines of the processes. (iv) Convoy Effect: - All the processes wait for one long process to get off CPU [11].

¹SGTBIM&IT, GGSIPU, Delhi

²Delhi Technological University, Delhi

E-mail: ¹pahwapre@gmail.com and ²ms2690@rediffmail.com

2. **SJF**: - To overcome the limitations of FCFS, Shortest Job First (SJF) algorithm was proposed. This algorithm selects the process with smallest burst time to execute next. The limitation of algorithm is: - it is very difficult to know the burst time of next CPU request. Although this algorithm is optimal but it cannot be implemented at the level of short-term CPU scheduling[11].
3. **SRTF**: - Shortest-Remaining-Time-First (SRTF) scheduling algorithm is preemptive version of SJF. This algorithm allows the next process with shorter burst to preempt the process already executing, if the burst of new arrived process is shorter than the remaining time for the running process.
4. **Priority Scheduling Algorithm (Pri)**: - In this algorithm the process with highest priority is assigned CPU first and so on. The priorities are assigned to process by operating system. Low priority process gets interrupted by the incoming of higher priority process. The limitation of algorithm is indefinite blocking or starvation of lower priority process if there is large number of high priority process. Also, waiting time and response time depends on priority of process. To overcome the limitation of indefinite blocking aging technique was proposed which gradually increases the priority of processes waiting from long time.

None of the algorithms stated above is ideal with respect to scheduling objectives. Therefore, in this paper we proposed a new algorithm which uses fuzzy logic to find the dynamic priority of the process.

2. RELATED WORK

Terry Regner & Craig Lacey[8] introduced the concepts and fundamentals of the structure and functionality of operating systems. The purpose of this article was to analyze different scheduling algorithms in a simulated system. This article has the implementation of three different scheduling algorithms: shortest process first, round robin, and priority sequence. Comparing the three algorithms, they find that the CPU utilization values indicate that the shortest process first has the highest throughput values with CPU utilization times comparable to those of the round robin. Ajit Singh[9] developed a new approach for round robin scheduling which helps to improve the CPU efficiency in real time and time sharing operating system. Alexander[10] stated that Multimedia applications have unique requirements that must be met by network and operating system components. In any multimedia application, we may have several processes running dependently on one another. Multimedia is a real-time

application. In context of multimedia applications, the CPU scheduler determines quality of service rendered. The more CPU cycles scheduled to a process, the more data can be produced faster, which results in a better quality, more reliable output. Many Researchers have tried to implement fuzzy logic to schedule the processes. A fuzzy-based CPU scheduling algorithm is proposed by Shata J. Kadhim et. al[1]. Round robin scheduling using neuro fuzzy approach is proposed by Mr. Jeegar A Trivedi et. al[2]. Soft real-time fuzzy task scheduling for multiprocessor systems is proposed by Mahdi Hamzeh et. al[3]. Efficient soft real-time processing is proposed by C. Lin et. al[4]. An Improved fuzzy-based CPU Scheduling(IFCS)algorithm for real time systems is proposed by H.S. Behera[5].

3. FUZZY LOGIC TERMINOLOGY USED

3.1 Fuzzy Logic:- A Fuzzy logic is a generalization of standard logic, in which a concept can possess a degree of truth anywhere between 0 and 1. It allows intermediate values to be defined between conventional evaluations. A Fuzzy logic system is nonlinear mapping of an input data to output data. A Fuzzy logic system consists of components: fuzzier, rules, inference engine and defuzzier. The process of fuzzy logic is to first collect the crisp set of inputs and convert it to the fuzzy sets using fuzzy linguistic variables, fuzzy linguistic terms and membership functions. This is known as Fuzzification. Afterwards an inference is made on the basis of set of rules. Finally, the resulting fuzzy output is mapped to a crisp output using the membership functions, in the defuzzification step.

3.2 Fuzzy Logic Terminology :-

(i) **Linguistic Variables:** - It is the input or output variables of the system whose values are non-numeric.

The values may be words or sentences derived from natural language.

(ii) **Membership Functions:** - Membership functions are used in the Fuzzification and defuzzification steps of a FLS, to map the non-fuzzy input values to fuzzy linguistic terms and vice-a-versa. A membership function is used to quantify a linguistic term. It is denoted by μ . In our proposed algorithm we consider two memberships, one of burst time (μ_b) and other of priority (μ_p).

3.3 Fuzzy Inference System (FIS):- Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic. The mapping then provides a basis from which decisions can be made[5]. An FIS consists of an input stage, a processing stage, and an output stage. The input stage maps the inputs, such as deadline, execution time, and so on, to the appropriate membership functions and truth values. There are two common FIS:-(i)Mamdani's fuzzy inference method proposed in 1975 by Ebrahim Mamdani[6]. (ii)Takagi-Sugeno-Kang, or simply Sugeno, method of fuzzy inference introduced in 1985[7].

These two methods are similar as the procedure of fuzzifying the inputs and fuzzy operators. The difference between the two is that the Sugeno's output membership functions are either linear or constant but Mamdani's inference expects the output membership functions to be fuzzy sets.

In our proposed algorithm we use Mamdani's inference system.

3.4 Fuzzy Based CPU Scheduling Algorithms:-

Improved Fuzzy CPU Scheduling Algorithm (IFCS):-

Scheduling is very critical for real time processes. The processes priorities keep on changing in real time. This algorithm is based on dynamic priorities (dpi) rather on static priorities (pti). It considers membership function of priority, burst time and response ratio to find dynamic priority and schedule the process according to it. The algorithm was proposed by H.S. Behera et al.[6]. This algorithm ensures reduced waiting time and turnaround time.

Proposed Fuzzy CPU Scheduling Algorithm (PFCS):-

This algorithm also calculates dynamic priorities and schedules the process according to it but it doesn't include membership function based on response ratio while calculating dynamic priority. For calculation of dynamic priorities it relies on membership function of priorities and burst time. The algorithm is further evaluated to see the performance in terms of turnaround time and waiting time.

4. SIMULATOR USED

4.1 PSSAV (Process Scheduling Simulation, Analyzer, and Visualization):-It is an application for CPU scheduling algorithm which provides customizable comparison between each scheduling algorithm. We have used to analyze our algorithm in this simulator.

4.2 Emulated turbo C++:-It is an integrated development environment (IDE) which has C compiler. We have developed the code corresponding to our algorithm in this IDE.

5. PROPOSED ALGORITHM (PFCS):-

5.1 Calculate dynamic priority(dpi):-

1) For each process P_i in ready queue fetch its parameters burst time (bti), static priority (pti), and arrival time (ati) and give them as input to FIS.

2) For each process P_i ; evaluate membership function of priority i.e. μ_p

$$\mu_p = \text{pti} / \max(\text{pti}) + 1; \text{ where } 1 \leq i \leq n$$

3) For each process P_i ; evaluate membership function of burst time i.e. μ_b

$$\mu_b = 1 - (\text{bti} / \max(\text{bti}) + 1); \text{ where } 1 \leq i \leq n$$

4) For each process P_i in ready queue find minimum priority process.

5) To calculate dynamic priority (dpi)

If process P_i has minimum priority then

$$\text{dpi} = (\mu_p + \mu_b)$$

Else

$$\text{dpi} = \max\{\mu_p, \mu_b\}$$

where $1 \leq i \leq n$

(A) Pseudo code:-

- 1) Set dynamic priority (dpi) to the output of FIS.
- 2) Schedule the process P_i with the highest value of dpi for execution where $1 \leq i \leq n$.
- 3) If the scheduled process finishes and no new process arrives go to step 2.
Else if new process arrives go to step 1.
- 4) Exit.

6. RESULTS & PERFORMANCE EVALUATION

Different Case Studies are being taken to evaluate the performance of proposed fuzzy CPU Scheduling Algorithm (PFCS).

CASE STUDY 1:-

Process ID	Arrival Time(ati)	Burst Time(bti)	Static Priority (pti)
P1	0	3	6
P2	2	6	1
P3	4	4	5
P4	6	5	2
P5	8	2	7

Table 1: Case Study-1 Data Set [5]

Comparison Table:-

Algorithm	Waiting Time average (ms)	Turnaround Time average (ms)	Completion Time (ms)
Priority Algorithm(Pri)	4.8	8.8	20
IFCS	3.8	7.8	20
PFCS	3.8	7.8	20

Table 2: Comparison between various algorithms for Case Study-1

CASE STUDY 2:-

Process ID	Arrival Time(ati)	Burst Time(bti)	Static Priority(pti)
P1	0	18	11
P2	0	2	6
P3	0	1	7
P4	0	4	3
P5	0	3	5
P6	0	12	2
P7	0	13	1

Table 3: Case Study-2 Data Set [5]

Comparison Table:-

Algorithm	Waiting Time average(ms)	Turnaround Time average(ms)	Completion Time(ms)
Priority Algorithm(Pri)	23.86	31.43	53
IFCS	14.86	22.43	53
PFCS	14.86	22.43	53

Table 4: Comparison between various algorithms for Case Study-2

Comparative Analysis:-

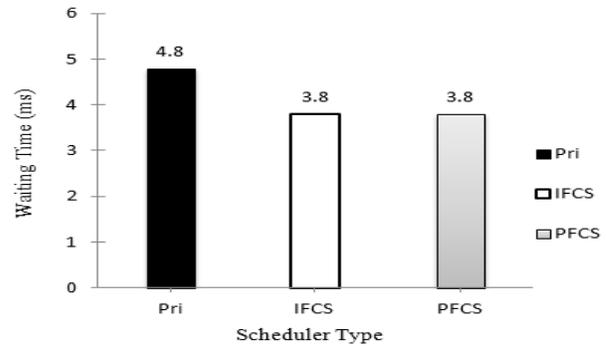


Figure 1: Waiting time comparison between various algorithms for Case Study-1.

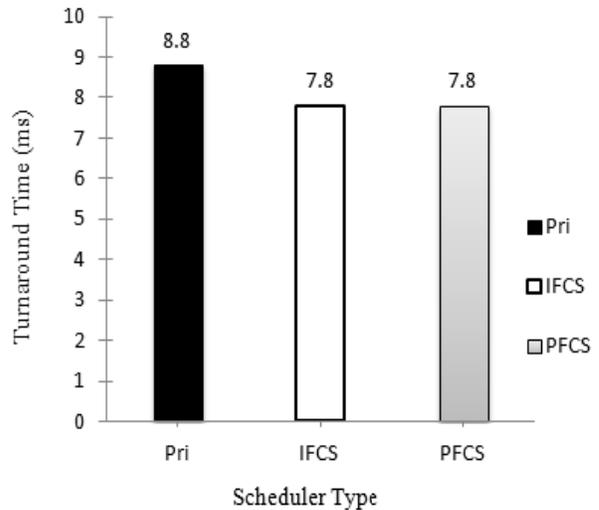


Figure 2: Turnaround time comparison between various algorithms for Case Study-1.

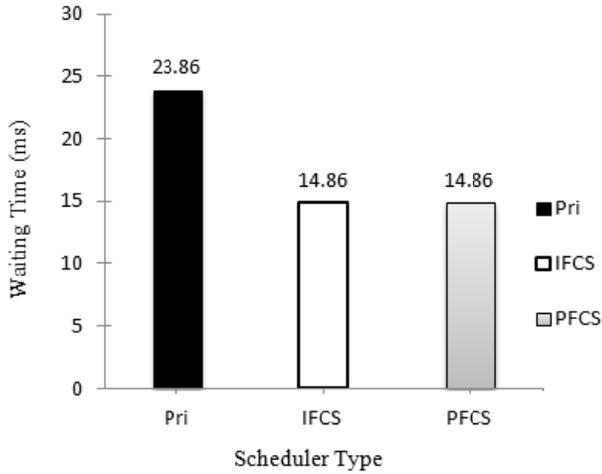


Figure 3: Waiting time comparison between various algorithms for Case Study-2.

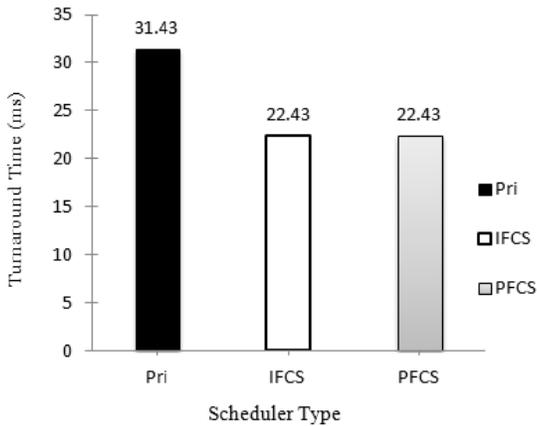


Figure 4: Turnaround time comparison between various algorithms for Case Study-2.

Comparison Table:-

Algorithm	Waiting Time average(ms)	Turnaround Time average(ms)	Completion Time(ms)
Priority Algorithm(Pri)	22.44	30.33	71
IFCS	20.78	28.67	71
PFCS	20.78	28.67	71

Table 5: Comparison between various algorithms for Case Study-3

Comparative Analysis:-

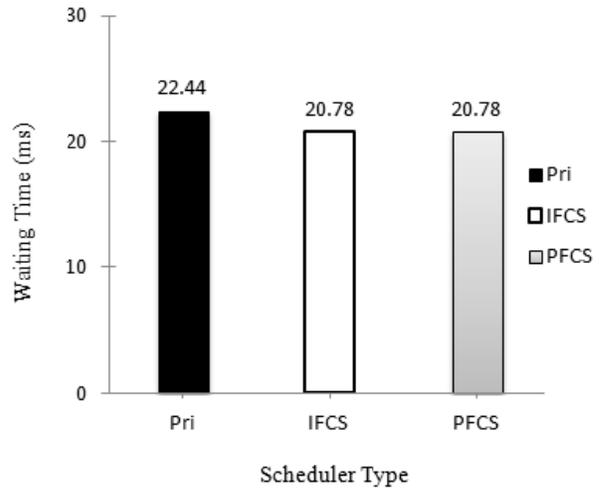


Figure 5: Waiting time comparison between various algorithms for Case Study-3.

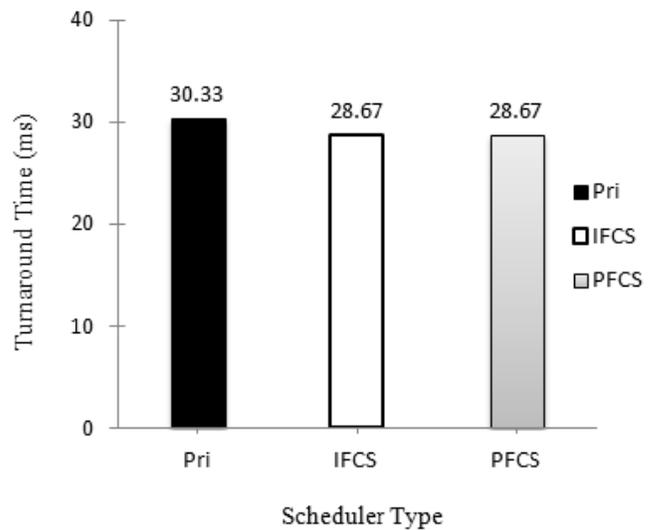


Figure 6: Turnaround time comparison between various algorithms for Case Study-3.

CASE STUDY 4:-

Process ID	Arrival Time(ati)	Burst Time(bti)	Static Priority (pti)
P1	0	10	3
P2	0	1	1
P3	0	2	4
P4	0	1	5
P5	0	5	2

Table 6: Case Study-4 Data Set [11]

Comparison Table:-

Algorithm	Waiting Time average(ms)	Turnaround Time average(ms)	Completion Time(ms)
Priority Algorithm(Pri)	8.2	12.0	19
IFCS	3.2	7.0	19
PFCS	3.2	7.0	19

Table 7: Comparison between various algorithms for Case Study-4

Comparative Analysis:-

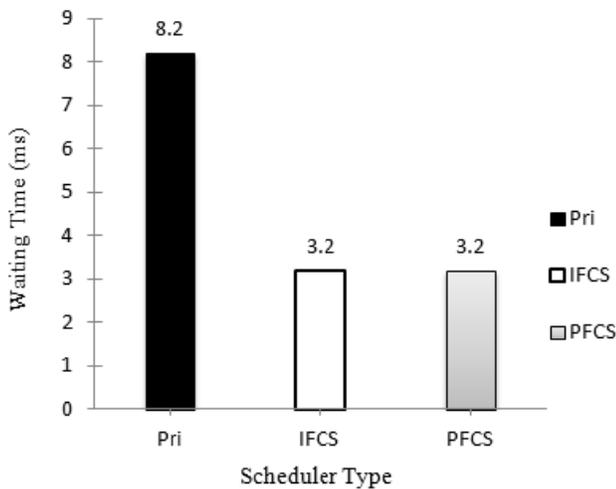


Figure 7: Waiting time comparison between various algorithms for Case Study-4.

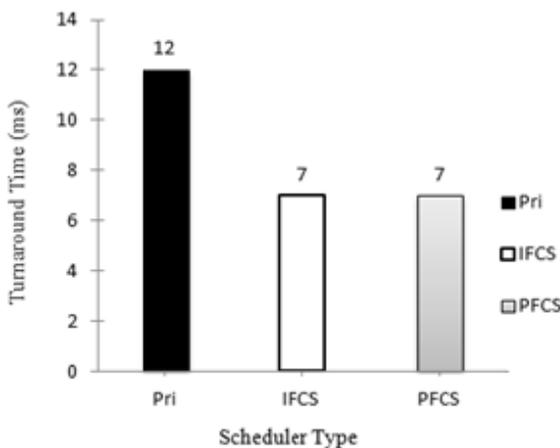


Figure 8: Turnaround time comparison between various algorithms for Case Study-4.

6.1 DISCUSSION ON SIMULATION RESULTS

It is clear from various case studies presented above that IFCS and PFCS show almost same result in terms of turnaround time and waiting time. These results show that system performance is far improved when compared to priority scheduling. Thus, when process is scheduled according to dynamic priority (dpi), it helps in reducing average turnaround time and average waiting time, so system performance is improved. But PFCS is more advantageous than IFCS as it reduces the extra burden of calculation of response ratio.

7. CONCLUSION AND FUTURE SCOPE

The Proposed Fuzzy CPU Scheduling Algorithm reduces cost of calculating response ratio. In order to obtain an efficient scheduling algorithm, two membership functions μ_p and μ_b deals with both task priority and its execution time are evaluated to find dynamic priority so as to overcome the shortcomings of well-known scheduling algorithms. This Proposed Fuzzy CPU Scheduling Algorithm can further be improved by improving Fuzzification process. A new Fuzzy neuro based CPU scheduling algorithm can be generated. This Proposed Fuzzy CPU Scheduling Algorithm can be further improved by choosing more accurate formula for evaluating fuzzy membership value which may further reduce the waiting time and turnaround time.

ACKNOWLEDGEMENT

First of all, let me thank the almighty GOD, my parents and my In- Laws who are the most graceful and merciful for their blessings that contributed to the successful completion of this research. I take the opportunity to offer sincere thanks and deep sense of gratitude to my learned supervisor Mr. Manoj Sethi for expressing his confidence in me. I am thankful to my husband Mr. Rahul Ajmani for his moral support all the time; he has been always around to cheer me up in the odd times.

REFERENCES

- [1]. Shata J. Kadhim, Kasim M. Al-Aubidy, "Design and Evaluation of a Fuzzy Based CPU scheduling Algorithm", *Springer-verlag Berlin Heidelberg 2010, CCIS 70*, pp. 45-52, 2010.
- [2]. Jeegar A Trivedi and Dr.Priti Srinivas Sajja, "Improving efficiency of round robin scheduling using neuro fuzzy approach", *IJRRCS*, Vol.2, No. 2, April 2011.
- [3]. Mahdi Hamzeh, Sied Mehdi Fakhraie and Caro Lucas, "Soft real time fuzzy task scheduling for multiprocessor systems", *International Journal of Intelligent Technology*, Vol.2, No. 4, 2007.
- [4]. C. Lin and S. A. Brandt, "Efficient soft real-time processing 1in an integrated system", in *Proc. 25th IEEE Real-Time S1ystems Symp*, 2004.
- [5]. H. S. Behera, Ratikanta Pattanayak and Priyabrata Mallick, "An Improved Fuzzy-Based CPU Scheduling (IFCS) Algorithm for Real Time Systems", *International Journal of Soft Computing and Engineering (IJSCE)*, Volume-2, Issue-1, March 2012.

- [6]. Mamdani E.H., Assilian S., “An experiment in linguistic synthesis with a fuzzy logic controller”, *International Journal of Man-Machine Studies*, Vol.7, No. 1, 1975.
- [7]. Sugeno, M., “Industrial applications of fuzzy control”, *Elsevier Science Inc.*, New York, NY, 1985.
- [8]. Terry Regner, craig lacy, “An Introductory Study of Scheduling Algorithms”, *CPSC 321: A Project Assignment*, Feb 2005.
- [9]. Ajit Singh, Priyanka Goyal, “An Optimized Round Robin Scheduling Algorithm for CPU Scheduling”, (*IJCSE*) *International Journal on Computer Science and Engineering*, Vol. 02, No. 07, 2010.
- [10]. Alexander Taranovsky, “CPU Scheduling in Multimedia Operating Systems”, *Research Report*, 1999.

Book References

- [11] Silberschatz, A., Peterson, J. L., and Galvin, P.B. - *Operating System Concepts*; 7th Edition; Addison Wesley, 2006.

Web Based Technologies to Support High Level Process Maturity

A. V. Sharmila

Submitted in March, 2013; Accepted in August, 2013

Abstract - This paper discusses the uses of Web based Technologies to support High Level Process Maturity in an organization. It also provides an overview of CMMI, focusing on the importance of centralized data storage and data access for sustaining high maturity levels of CMMI. Further, elaboration is made on the web based technology, stressing that change over to Web Based Application is extremely helpful to maintain the centralized data repository, to collect data for process capability baseline, and to track process performance management, with reduced maintenance effort and ease of data access. A case study analysis of advantages of adopting Web Based Technology is also narrated. Finally the paper concludes that the sustenance of High level Process maturity can be achieved by adopting web application technology.

Index Terms - CMMI, management of organizational process improvement, centralized repository, web technology, effort booking and automated web based tools.

1. INTRODUCTION

The Intranet and Internet-based technologies are impacting businesses in many ways. New technologies are promising to save costs, to improve customer and supplier relationships, business processes, and performance, and to open new business opportunities. They allow firms to respond better to existing challenges and improve the anticipation of future developments. While traditional use of information technology (IT) has been mostly at the back-end, and for applications supporting production-oriented procurement and supply-chains of large companies (the traditional area of EDI-systems), a recent trend is focusing on user-friendly front-end applications supporting non-production-oriented procurement by non-purchasing experts. The numbers of operational customer implementations are growing fast. Individual attempts to connect inter-organizational business processes and to interchange information are often undertaken independently from each other, and, due to a lack of commonly accepted standards, leading to incompatible results [1].

This paper addresses the use of web based technologies to support non production oriented processes. We discuss the implications of web based systems on High Level Process maturity and some of the critical success factors and key decision points that firms have to consider in order to utilize the technologies in an optimal way.

E-mail: avsharmila@bel.co.in

An example from the BEL Software Technology Center is showcased to clarify the concepts. By presenting results from case and field study research, we address researchers as well as business managers who want to understand how emerging technologies are supporting High Level Process Maturity.

2. BACKGROUND FOR THE NEED

The BEL Software Technology Centre (BSTC), a leading Defence Public Sector Unit (DPSU), has followed the staged representation of CMMI. The organization was successfully appraised against SW CMM Level 4 in September 2004, Maturity Level 5 of the CMMI Version 1.2 in June 2009 and recently appraised for sustenance of SEI-CMMI L5Version 1.3 in May 2012.

The CMMI model, developed by a group of industry, government, and SEI representatives, is made up of best-of-the-best processes gleaned from multiple disciplines. It provides guidance in specific process areas through a set of goals and expected practices needed to meet those goals. The BEL Software Technology Centre has successfully adopted the CMMI best practices and follows statistically measured and controlled software process improvements [2].

For the Effective Management of Organizational Process Improvements the following data are required from Engineers: Engineer's Skill Set, Project & Task Allotment, Effort Hour accounted against project tasks, Suggestions for improvements, Root Cause analysis for Problems Faced, Trainings required and Attended. The organization has more than a couple of hundreds of engineers, working in various buildings, and connected through the organization's Intranet. This Situation pushed us to think about a centralized repository to capture and access the data from a Single place, with ease of access and maintenance. The Web Based application has a Client-Server Architecture, providing the single server-multiple clients concept. Thus it was decided to develop an Automated Web Based in House Tool to Support High Level Process maturity.

3. WEB BASED TECHNOLOGY

3.1 Overview

One of the early applications of the Internet was its most popular application, the World Wide Web or www. The www is one of the software tools that, through the use of hypertext, allow computers to link information in new ways different from a sequential reading approach, to make it easy to retrieve and add information from different computer sources through the use of communication links [3].

In a short time since its inception, the Internet has indeed revolutionized business, in that it redefines the methods used in traditional business practices and offers another important

channel for mass communication[4].Initially, the Internet was primarily utilized as a medium for communication (E.g. e-mail) purposes. Soon after many organizations from both public and private sectors began to discover that they could utilize this technology in support of marketing and information dissemination purposes. This resulted in organizations realizing that the greatest payback in investing in the technologies of www would be sharing information about the organization's products and services with the organization's stakeholders [5]. As a result, successful organizations of all sizes and types have been adopting different applications/technologies of www and discovering emerging ways of doing business that even a decade ago could not be imagined [6].

3.2 AVAILABLE TYPES

There are three main types of web applications:

- **Customer-facing applications** are known as ecommerce or B2C sites and use the internet. These typically present a customer with choices of products or services to buy using a shopping cart and payment method.
- **Employee-facing applications** use the intranet in a company. One example is a company's Accounting application. Another might be employee expense reporting. These applications previously operated on an internal client-server network. They are now web-enabled to make them easier to use and deploy.
- **Customer-Supplier facing applications** are known as B2B (Business to Business) sites and use the extranet, (an extension of an intranet that allows outside companies to work in a password-protected space). B2B sites provide a secure means for sharing selected information.

Not all web applications fit the above categories. For example Yahoo! email is not in any of the above. However, the above categories are representative of the main types of applications.

3.3 ARCHITECTURE

Web applications are built using a 3-tier architecture in which the client, server and database constitute the main elements [7]. The basic structure of a 3-tier architecture is indicated in the figure 1:

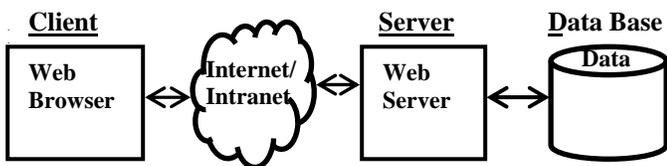


Figure 1: Basic structure of a 3-tier application architecture

3.4 TECHNOLOGIES USED TO BUILD WEB APPLICATION

There are 6 competing technology stacks that differ in the following attributes [8]:

- **Programming languages (Lang)** - can be used to create programs that control the behavior of a machine and/or to express algorithms precisely.
- **Operating system (OS)** is a collection of software that manages computer hardware resources and provides common services for computer programs. The operating system is a vital component of the system software in a computer system. This can be Linux (L), Unix (U) or Windows (W).

Web server (Server) - can refer to either the hardware (the computer) or the software (the computer application) that helps to deliver Web content that can be accessed through the Internet/Intranet.

Database support (DB) – A Database Server is a computer program that provides database services to other computer programs or computers, as defined by the client–server model.

Sponsoring companies (Sponsors) - A sponsor is the individual or group that provides the support, similar to a benefactor.

The following table summarizes these technology stacks.

Stack	Sponsor	OS	Server	DB	Lang
CGI	Open source	L/U	Apache	Varies	Perl
Cold fusion	Macro media	W/L/U	Cold fusion	Varies	CFML
LAMP	Open source	L/W/U	Apache	MySQL	PHP
Java/J2EE	Sun, IBM	L/U	J2EE	Varies	Java
.NET	Microsoft	W	ASP.NET	SQL server	VBasic C#
Java/J2EE	Open Source	L/U/W	Any Java Web Server	Varies	Java

Table 1: Technology Stacks

Adopted Type and Stacks: Since our Engineers are connected through intranet, it is identified as the best suitable web application type is employee-facing application. The adopted stacks for the in house developed tools in BEL Software Technology Center are

For New Tools: SQL Server (DB), Apache Tomcat Server (Web Server), OS (L/U/W), Language (Java).

For Existing Tool (Developed in VB): SQL Server (DB), ASP .NET (Server), OS (W), Language (C#).

4. CASE STUDY

4.1 Establishment of Effort Capturing Web Based Tool in Bel Software Technology Centre

Need: It is required to track the effort hours accounted in the organization, with respect to projects and supporting tasks (such as quality assurance, testing, configuration management, network administration etc), performed by each of the engineers. This is needed to be accounted weekly, and collated on monthly basis as the total effort expended under each of the work orders executed in the organization.

4.2 Why Effort Booking is Important?

Effort data collected accurately helps to track projects and supporting tasks with respect to planned effort for the project or activity. The effort history available consistently over a period of time supports the estimation of effort hours required to perform future projects and supporting tasks.

4.3 Challenge

The organization has various groups of engineers, spread out over different facilities / buildings of the organization. They all expend effort, contributing to a set of projects and supporting tasks of the organization. Engineers working on the same project or same supporting function are not strictly co-located at the same facility. From all the facilities, the project managers are needed to allot tasks to the engineers and the engineers are required to book their expended effort. From the common repository, the effort needs to be collated and status / tracking reports as well as history data have to be generated. The situation is depicted pictorially in figure 2.

Solution: A Web Based Tool is developed in BEL Software Technology Center to log the project details, engineer details, and effort hours spent towards software development and support activities. The tool maintains the data in a common repository.

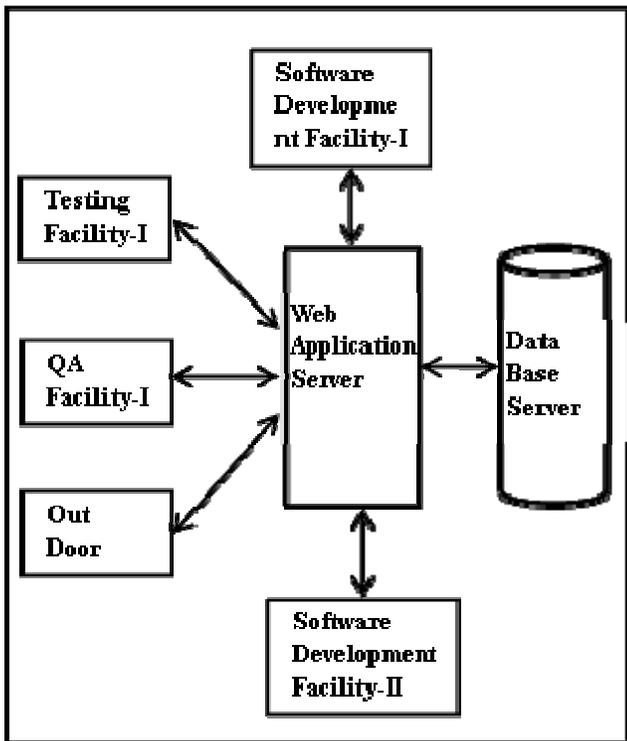


Figure 2: Multiple groups accessing the common data from different facilities

4.3.1 Approach

Task List: Detailed Task List was identified for different activities performed during the software development and support tasks.

Installation and Configuration: SQL server was installed on the development machine and Server machine. IIS was installed and configured to deploy the packages.

Package1- Engineer WRS: Consists of activity task list and enables the week wise effort hour booking for the engineers.

Package2- Manager WRS: Provides options to enter project details, engineer details and task assignment. Provides report generation on booked engineer’s effort hours.

Deployment: The package is deployed from the server and the same is accessed through the web browser for all machines connected to the organization’s Intranet. Browsers supported are IE6 or above and Firefox 3 or above.

Benefits:

1. Centralized Repository is maintained, leading to ease of data collection and improved process performance management.
2. Tool is installed centrally on a single server, avoiding the effort, time and space required to install in individual engineer’s PCs.
3. New releases of the Software are installed at central server, eliminating the need to update each work station. This significantly reduces the on-going maintenance costs.
4. Network traffic is minimized because users of the system receive only the information they need.
5. Training Cost is reduced because end users are already familiar with the functionality of their browsers and web application in general. So it is easier for them to adapt to web based systems.

5. CONCLUSION

The Web Based Technology is offering new emerging capabilities for sharing and disseminating information. In this paper, we provided an overview of how web based technologies can support High Level Process maturity. A case study on Effort Booking Server as showcases to illustrate the issues. A Cost Effective Process Performance Management can be achieved by adopting Web based Technology. This in turn boosts the organization’s ability to sustain the operations at the high level process maturity with ease.

FUTURE SCOPE

In BEL Software Technology Center, currently the in-house developed tools are used for various process automation aspects such as suggestion management, training management, effort and task management, decision and causal analysis records and knowledge management. These tools are based on common standardized web technologies and are being maintained by in-house development teams. The future work would involve activities for integrating these tools and maintaining a common catalogued centralized repository, further easing the maintainability.

REFERENCES

[1]. Berners-Lee, Cailliau & Groff, 1992. Berners-Lee, T.J., Calilliau and Groff, J.F. (1992). The World Wide

- Web.Computer Networks & ISDN Systems, 25(4), 454-459.
- [2]. BSTC Process Document, (BSTC-PEG, 2012).
 - [3]. Foo, S. and Lim, E.(1997). A Hypermedia Database to Manage World-Wide Web documents. Information and Management, 31(5), 235-249.
 - [4]. Khosrow-Pour, M. and Herman, N. (2001). "Web-Enabled Technologies Assessment and Management: Critical Issues, Managing Web-Enabled Technologies in Organizations: A Global Perspective, Mehdi.
 - [5]. Khosrow-Pour and Nancy Herman (eds.), Hershey, PA: Idea Group Publishing, 1-22Prawitt, Romney & Marshal, 1997; Khosrow-Pour & Herman, 2000)
 - [6]. Troller, Michael. "How the Internet Works".(c)2002 Que.
 - [7]. Vermeulen, Bert. The Corp21.com site as several related articles on topics such as "Making Your Website Popular", "Promoting your website Through Search Engines" and "Web Basics."
 - [8]. Zhao, Jenson J. "Web Design and Development for eBusiness". (c)2003 Prentice Hall.

Energy Payback Time Calculation for a Building Integrated Semitransparent Thermal (BISPVT) System with Air Duct

Kanchan Mudgil¹ and Deepali Kamthania²

Submitted in April, 2013; Accepted in August, 2013

Abstract - This paper evaluates the energy payback time (EPBT) of building integrated photovoltaic thermal (BISPVT) system for Srinagar, India. Three different photovoltaic (PV) modules namely mono crystalline silicon (m-Si), poly crystalline silicon (p-Si), and amorphous silicon (a-Si) have been considered for calculation of EPBT. It is found that, the EPBT is lowest in m-Si. Hence, integration of m-Si PV modules on the roof of a room is economical.

Index Terms - Embodied energy, Energy payback time, Photovoltaic

1. INTRODUCTION

Development in the design and manufacture of photovoltaic cells has been very rapid over the last few years because they are now predicted to become a major renewable energy source. The embodied energy payback time is important for renewable technologies as their use makes no sense if the energy used in their manufacture is more than they can save in their life-time. Energy payback time (EPBT) of a PV module is defined as the ratio of energy consumed to produce the PV system to annual energy generated from the PV system by using solar energy. Slesser and Houman [1] have been reported that the energy payback time (EPBT) of a PV module is 40 years. Aulich et al. [2] have evaluated the energy payback time of 8 years for a crystalline silicon module. The energy payback time for a crystalline silicon solar cell module under Indian climatic condition for annual peak load duration is about 4 years, Prakash and Bansal [3]. Keolein and Lewis [4] have predicted the energy payback time (EPBT) of 7.4 years for an amorphous silicon (a-Si) solar cell module for the climatic conditions of Detroit, USA.

Yamada et al. [5] have evaluated the energy payback time (EPBT) for both polycrystalline and amorphous silicon solar cell and reported the energy payback time of 5.7 and 6.3 years respectively at annual power production rate of 0.01 GW/y. Battisti and Corrado [6] have investigated the energy payback time (EPBT) for a conventional multi-crystalline building integrated system, retrofitted on an inclined roof, located in Rome (Italy); with yearly global insolation on a horizontal

plane was taken as 1530 kWh/m²y. They have concluded that energy payback time (EPBT) gets reduced from 3.3 years to 2.8 years. In this paper, energy payback time (EPBT) based on overall thermal energy and exergy output from BISPVT system with duct, for Srinagar, India has been evaluated.

2. EMBODIED ENERGY

Embodied energy is defined as: "the quantity of energy required by all of the activities associated with a production process, including the relative proportions consumed in all activities upstream to the acquisition of natural resources and the share of energy used in making equipments and in other supporting functions i.e. direct energy plus indirect energy",

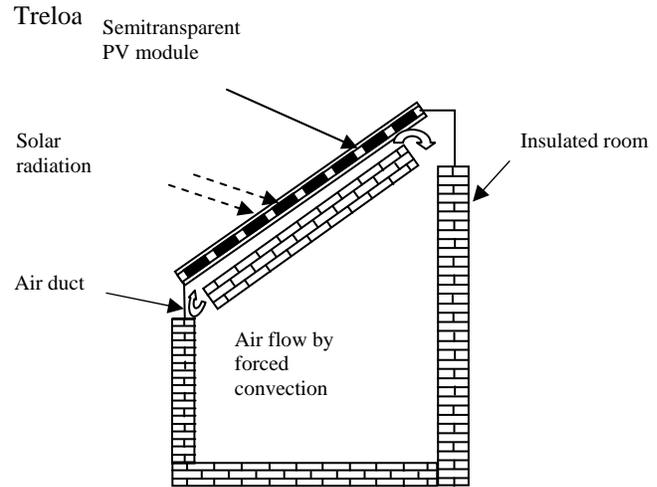


Figure 1: Cross section view of building integrated photovoltaic thermal (BISPVT) system integrated to roof with air duct.

Thus embodied energy analysis is quantifying the amount of energy used to manufacture a material or component. This involves the assessment of the overall expenditure of energy required to extract the raw material, manufacture product or components, installation and maintain the component element whichever is being assessed. For the embodied energy analysis of BISPVT system integrated to roof with air duct (shown in Fig. 1), the total energy requirement for individual components with their manufacturing energy needs to be evaluated. The break up of embodied energy of each component of BISPVT system has been tabulated in Table 1. BISPVT system consists of building materials, DC fan, different PV materials, battery, inverter, PV frame and charge controller.

¹Maharaja Agrasen institute of Technology, Sector-22, Rohini, New Delhi

²Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), A-4 Paschim Vihar, New Delhi
E-mail: ²deepali102@gmail.com

3. ENERGY PAY BACK TIME (EPBT)

The EPBT depends upon the energy spent to prepare the materials used for fabrication of the system and its components, i.e. embodied energy and the annual energy yield (output) obtained from such systems. To evaluate embodied energy of various components of systems, the energy densities of different materials are required. It is the total time period required to recover the total energy spent to prepare the materials (embodied energy) used for fabrication of BISPVT system. It is defined as the time necessary for a system to generate the energy equivalent to that used to produce, install and decommission. Mathematically,

$$EPBT = \frac{\text{Total energy input in manufacturing, installation and decommission}}{\text{Annual energy output from the system}} = \frac{E_{in}}{E_{out}} \quad (1)$$

$$T_{epb} = \frac{\text{Embodied energy } (E_{in})}{\text{Annual energy output } (E_{out})} \quad (2)$$

EPBT for the BISPVT system with air duct can be expressed as,

$$EPBT = \frac{E_{building} + E_{support} + E_{BISPVT} + E_{inverter} + E_{inst+M\&O} + E_{dec}}{E_{aout}} \quad (3)$$

where, $E_{building}$, $E_{support}$, E_{BISPVT} , $E_{inverter}$, $E_{inst+M\&O}$, E_{dec} are the embodied energy of the building, support for the BISPVT system, BISPVT system, inverter, battery, installation maintenance and operation and decommissioning respectively.

Building construction material	Specific energy content (MJ/ kg)		Quantity used	Embodied energy (MJ)
	Range	Assigned		
Clay bricks (23×11×7.5 cm ³) (for foundation)	2-7	4.5	1632 kg (510 Nos.)	7344
Clay bricks (23×11×7.5 cm ³) (for floor)	2-7	4.5	1600 kg (500 Nos.)	7200
Clay bricks (23×11×7.5 cm ³) (for walls)	2-7	4.5	4896 kg (1530 Nos.)	22032
Cement	4-8	6	750 kg	4500
sand	< 0.5	0.2	129.2 kg	129.2
concrete (for foundation, roof and walls)	0.8-1.5	1.15	646 kg 1294 kg	1488.1
Lime (for white wash)	3-5	4	0.5 kg	2
Mild steel (for roof)	30-60	45	78 kg	3510

Building construction material	Specific energy content (MJ/ kg)		Quantity used	Embodied energy (MJ)
	Range	Assigned		
Paint (for doors)	80-150	115	0.5 kg	57.5
Plywood (for one door)	8-12	10	9 kg	90
Straw fibre		0.28	5 kg	1.45
				46354.15 MJ
				12876.15 kWh

Table 1(a). Embodied energy of the building materials used for BISPVT system with air duct (Agrawal [8]).

DC fan (Exhaust fan)	Specific energy content (kWh/ kg)	Quantity used	Embodied energy (kWh)
Aluminium	32.39	0.39 kg	12.63
Iron	8.89	0.22 kg	1.95
Plastic	19.44	0.12 kg	2.33
Copper wire	19.61	0.05 kg	0.98
Fittings			
Hinges	32.39	0.2 kg	6.47
Door lock	32.39	0.025 kg	0.80
Hooks	32.39	0.1 kg	3.23
Nut and Bolt with washer, steel, screw & rivets	9.67	1.0 kg	9.67
			38.09 kWh

Table 1 (b): Embodied energy used for operating the DC fan for forced circulation of air used in BISPVT system with duct (Tiwari and Barnwal [9]).

Type of PV module	Expected life	Specific energy content (kWh/ m ²)		Quantity (No.)	Area of PV (m ²)	Embodied energy (kWh)
		Range	Assigned			
m-Si	30	1120-1260	1190	9	0.605	6479.55
p-Si	30	840-980	910	9	0.605	4954.95
a-Si	20	308-448	378	9	0.605	2058.21

Table 1 (c): Embodied energy for different PV materials (Agrawal [8]).

	Specific energy content, (kWh)	Quantity used	Embodied energy (kWh)
Battery	46	6	276
Inverter	1703	1	1703
M&O	59.5	-	59.5
PV frame	32.39 kWh/kg	4.5 kg	145.75
Charge controller	100	1	100
Miscellaneous	200	-	200
			2484.25

Table 1 (d): Embodied energy for battery, inverter, PV frame and charge controller. (Agrawal [8]).

4. RESULTS AND DISCUSSION

Annual overall thermal energy and exergy gain of BISPVT system for different PV material for Srinagar, India (considering a, b, c, and d type weather) have been shown in Fig. 2. The basic heat transfer equations derived by Vats et al. [10] have been used to compute the overall energy and exergy for different type of PV materials.

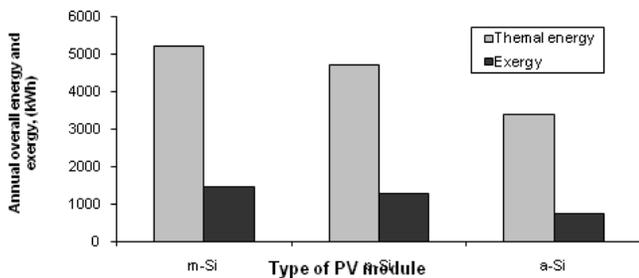


Figure 2: Type of PV Module

The values of energy and exergy observed from the Fig. 2 for BISPVT system with air duct have been tabulated in Table 2. The annual solar radiation (E_{sol}) falling on inclined roof, which is equal to the latitude of Srinagar (34 °) is calculated as 10465 kWh.

S. No.	Different types of PV materials	Annual overall thermal energy output (kWh),	Annual overall exergy output (kWh), $E_{aout,ex}$
1.	m-Si	5204	1479
2.	p-Si	4715	1285
3.	a-Si	3393	762

Table 2: Annual overall thermal energy and overall exergy for different type of PV materials, for a BISPVT system integrated to roof with duct.

The calculated values of EPBT for BISPVT system on energy and exergy basis have been shown in Table 3. From the table, it

is found that the energy payback time in BISPVT system integrated to roof with air duct is lowest in m-Si on thermal (4.26 years) and exergy (15.00 years) basis. Hence, EPBT point of view, use of m-Si in BISPVT system with duct is most suitable.

S. No.	Different types of PV materials	Embodied Energy (kWh), (E_{in})	Annual overall thermal energy output (kWh) $E_{aout,th}$	Annual overall exergy output (kWh), $E_{aout,ex}$	(T_{epb}) _{energy} (yrs.)	(T_{epb}) _{exergy} (yrs.)
1.	m-Si	21877	5204	1479	4.26	15.00
2.	p-Si	20352	4715	1285	4.38	16.00
3.	a-Si	17456	3393	762	5.26	23.00

Table 3: Calculations of energy payback time for different types of PV materials, for a BISPVT system integrated to roof with duct.

5. CONCLUSION

1. EPBT is lowest in m-Si PV modules integration as compared to p-Si and a-Si PV modules integration.
2. Use of m-Si PV on BISPVT system is economical as EPBT is lowest.

REFERENCES

- [1]. M. Slesser, I. Hounam, Solar Energy Breeders, Nature 262, 244, 1976.
- [2]. H. A. Aulich, F. W. Schulz, B. Strake, "Energy payback time for crystalline silicon photovoltaic modules using new technologies", *IEEE PV Specialists Conference*, Piscataway, NJ, 1213-1218, 1986.
- [3]. R. Prakash, N. K. Bansal, "Energy analysis of solar photovoltaic module production in India", *Energy Sources*, vol. 17, 605-613, 1995.
- [4]. G. A. Keolein, G. McD Lewis, "Application of life cycle energy analysis to PV module design", *Progress in photovoltaic: Research and Applications*, vol. 5, 287-300, 1997.
- [5]. K. Yamada, H. Komiyama, K. Kato, A. Inaba, "Evaluation of photovoltaic energy systems in terms of economics, energy and CO₂ emissions", *Energy Conversion Management*, vol. 36 (6-9), 819-822., 1995
- [6]. R. Battisti, A. Corrado, "Evaluation of technical improvements of photovoltaic system through life cycle assessment methodology", *Energy*, vol. 30, 952-967, 2005.

- [7]. G. J. Treloar, "Energy analysis of the construction of office buildings". Master of Architecture thesis. Deakin University, Geelong, 1994.
- [8]. B. Agrawal, "Performance evaluation of a building integrated photovoltaic thermal (BIPVT) system", PhD Thesis. Centre for Energy Studies, IIT New Delhi, India, 2010.
- [9]. G. N. Tiwari and P. Barnwal -Fundamentals of solar dryers. Anamaya Publishers, New Delhi, 2008.
- [10]. K. Vats, V. Tomar, G. N. Tiwari, "Effect of packing factor on the performance of a building integrated semitransparent photovoltaic thermal (BISPVT) system with air duct", *Energy and Buildings*, vol. 53, 159-165, 2012.

Effects of Using Filter Based Feature Selection on the Performance of Machine Learners Using Different Datasets

Mehnaz Khan¹ and S. M. K. Quadri²

Submitted in April, 2013; Accepted in August, 2013

Abstract - Data preprocessing is a very important task in machine learning applications. It includes the methods of data cleaning, normalization, integration, transformation, reduction, feature extraction and selection. Feature selection is the technique for selecting smaller feature subsets from the superset of original features/attributes in order to avoid irrelevant and additional features/attributes in the dataset and hence increases the accuracy rate of machine learning algorithms. However, the problem exists when the further removal of such features results in the decrease of the accuracy rate. Therefore, we need to find an optimal subset of features that is neither too large nor too small from the superset of original features. This paper reviews different feature selection methods- filter, wrapper and embedded, that help in selecting the optimal feature subsets. Further, the paper shows effects of feature selection on different machine learning algorithms- NaiveBayes, RandomForest and kNN). The results have shown different effects on the accuracy rates while selecting the features at different margins.

Index Terms - Data preprocessing, feature extraction, feature selection, dataset.

1. INTRODUCTION

In machine learning applications one of the most important tasks is data preprocessing [1]. The data that are collected for training in the machine learning tasks are not appropriate for the training purposes initially. In order to make the data useful for such applications, it needs to be processed. Processing involves methods for handling missing data [2] and methods for detecting and handling noise [3]. Data preprocessing is performed in order to prepare the data for input into machine learning and mining processes. This involves transforming the data for improving its quality and hence the performance of the machine learning algorithms, such as predictive accuracy and reducing the learning time. At the end of the data preprocessing stage, we get our final training set. One of the tasks of data preprocessing is feature selection in which only some of the features from the dataset are selected and used in the training process of the learning algorithm.

^{1,2}Department of Computer Science, University of Kashmir, India
E-mail: mhnzkhan6@gmail.com

In this process the aim is to find the optimal subset that increases the efficiency of the learning algorithm. Features in a dataset can be relevant i.e. the features that have influence on the output or irrelevant i.e. the features that have no effect on the output. Thus feature selection involves identifying the relevant features and using them in the machine learning application and ignoring the rest of the features with little or no predictive information.

The most important purpose of feature selection is to make a classifier more efficient by decreasing the size of the dataset. This is necessary for the classifiers that are costly to train e.g. NaiveBayes. The processing time and the cost of the classification systems are increased while their accuracy is decreased if irrelevant and additional features are used in the datasets being used for classification. Therefore, it is very important to develop the techniques for selecting smaller feature subsets. However, it has to be made sure that the subset which is selected is not so small that the accuracy rates are reduced and the results lack understandability. So it is very important that techniques must be developed that help to find an optimal subset of features from the superset of original features.

Feature selection comes with two approaches. One is called *forward selection* in which the process starts with no attributes/features which are then added one by one. At each step, the feature that decreases the error the most is added and the process continues until the addition of the features does not significantly decrease the error. Second approach is called *backward selection* in which the idea is to start with all the attributes/features and then remove them one by one. The feature to be removed at each step is the one that decreases the error the most, and the process is carried on until any further removal increases the error significantly.

In Section 3, different feature selection methods- filter, wrapper and embedded, that help in selecting the optimal feature subsets have been explained. Section 4 lists the basic steps that have been used for feature selection. Further, Section 5 gives the details of the experiment that was carried out using different machine learning algorithms on real data sets- Australian Credit Approval dataset from UCI Repository of Machine Learning Databases and Domain theories, Congressional Voting Records Dataset, and Adult Dataset. The optimal feature subsets achieved from the experiments have been explained in results and conclusion section.

2. RELATED WORK

This section presents the work done in the field of feature selection. A method of feature selection, called RELIEF has

been given that assigns a relevance weight to each attribute using instance based learning [4]. A book on feature selection has been given that includes all the feature selection methods that have been developed since 1970s and also gives a framework that helps in studying these methods [5]. Wrappers for feature subset selection have been developed in which an optimal feature subset is searched that is tailored to a particular learning algorithm and a particular training set [6]. The FOCUS algorithm has been designed for noise-free Boolean domains and it follows the MIN-FEATURES bias. It examines all feature subsets and selects the minimal subset of features that is sufficient to predict the class targets for all records in the training set [7]. Information gain and gain ratio are good examples of measuring the relevance of features for decision tree induction. They use the entropy measure to rank the features based on the information gained; the higher the gain the better the feature [8]. A feature selection model has been proposed using an instance-based algorithm, called RACE, as the induction engine, and leave-one-out cross-validation (LOOCV) as the subset evaluation function [9]. Emphasis has been laid on the issues of irrelevant features and the subset selection. It has been concluded that features that are selected should be dependent on the features and the target concept, as well as on the induction algorithm [10]. The forward and backward stepwise methods on the Calendar Apprentice domain have been designed, using the wrapper model and a variant of ID3 as the induction engine [11]. A method of feature selection for SVMs has been developed. The idea behind this method is to find those features which minimize bounds on the leave-one-out error. They have shown the method to be efficient as compared to some standard feature selection algorithms by testing on the datasets [12]. Twelve feature selection methods have been compared and a new feature selection metric called bi-normal separation (BNS) has been shown [13]. An introduction to variable and feature selection has been given that has suggested the use of a linear predictor e.g. a linear SVM and selection of variables in one of the two alternate ways. One is to use a variable ranking method using a correlation coefficient or mutual information and the other with a nested subset selection method that performs forward or backward selection [14]. A survey of feature selection methods for classification has been given [15]. A comparative study of feature selection methods in statistical learning of text categorization has been given that has evaluated document frequency (DF), information gain (IG), mutual information (MI) [16].

3. METHODS OF FEATURE SELECTION

Feature selection is regarded as a search problem in a space of feature subsets for which we need to specify a starting point, a strategy to traverse the space of subsets, an evaluation function and a stopping criterion [17]. There are three ways in which feature selection can be carried out. These are the filter, wrapper and embedded approaches. These methods differ in how they combine feature selection search with the construction of classification model [18, 23].

3.1 Filter Method

The filter approach selects a subset of the features that preserves as much as possible the relevant information found in the entire set of features. The methods that use the filter approach are independent of any particular algorithm as the function that they use for evaluation relies completely on properties of the data [24]. The relevance of the features is calculated by considering the intrinsic properties of the data. This involves the calculation of a feature relevance score and the features whose score is less are removed and only the remaining subset of features are used as input to the algorithm. Some filter methods use correlation coefficients like that of Fisher's discriminant criterion. Other methods use mutual information or statistical tests. Initially the filter-based methods did not take into consideration the relations between features but calculated the relevance of each feature in isolation. However, now the filter methods take many criteria into consideration e.g. now the filter methods select features with minimum redundancy. The most important feature selection framework used by many filter methods is the minimum-redundancy-maximum relevance (MRMR) framework [28]. Further the filter methods can be univariate or multivariate. Univariate filter methods take into account only one feature's contribution to the class at a time, e.g. information gain, chi-square. These methods are computationally efficient and parallelable however they are likely to select low quality feature subsets. On the other hand, multivariate filter methods take the contribution of a set of features to the class variable into account at a time, e.g. correlation feature selection and fast correlation-based filter. These methods are computationally efficient and select high quality feature subsets than univariate filters.

3.1.1 Advantages of Filter Approach

Some of the advantages of filter approach are:

- Filter methods of feature selection can be easily scaled to very high-dimensional datasets.
- These methods perform very fast and are computationally simple.
- They are not dependent on any particular algorithm.
- In these methods, feature selection is to be carried out only once, and then different classifiers can be evaluated.
- These methods have better computational complexity as compared to the wrapper methods.

3.1.2 Disadvantages of Filter Approach

Filter approach has some drawbacks:

- These methods do not take into account the interaction with the classifier. In other words, this method separates the search in the feature subset space from the search in the hypothesis space.
- In this method each feature is measured separately and thus does not take into account the feature dependencies.
- Lack of feature dependencies results in the degraded performance as compared to other techniques.

However, this problem is solved by a number of multivariate filter techniques that involve feature dependencies to some extent.

Examples of this approach: Euclidian distance, t-test, information gain, correlation based feature selection, Markov blanket filter.

3.2 Wrapper Method

Filter methods use a function for evaluation that relies on the properties of data and thus is not dependent on any algorithm. On the other hand, wrapper methods make use of the inductive algorithm for calculating the value of a given subset. These methods take into account the biases of the algorithm and thus are considered to be a better alternative in supervised learning problems. Wrapper methods [27] include the model hypothesis search within the feature subset search. In this method, the subsets of features are selected by first defining a search process in the possible feature subsets space, followed by generating and evaluating various subsets of features. The subsets of features are evaluated by training a specific classification model. This makes the wrapper method algorithm specific. Then for searching the space of all possible feature subsets, this method wraps a search algorithm around the classification model. But the search process requires a number of executions which results in a high computational cost, especially when more extensive search strategies are used. The search methods are of two types: deterministic and randomized search algorithms. Wrapper methods make use of the classifier for scoring the subsets of features based on their predictive power. A number of wrapper methods have been developed that are based on SVM. Support Vector Machine Recursive Feature Elimination is a wrapper method that makes use of a backward feature elimination scheme for eliminating insignificant features from subsets of features. In this method the features are ranked on the basis of the amount of reduction in the function. The feature selected for elimination is the one with the lowest rank.

3.2.1 Advantages of Wrapper Method

Advantages of wrapper approaches are:

- These methods involve the interaction between feature subset search and model selection.
- Wrapper methods take into account feature dependencies.
- Implementing a wrapper method is quite easy and straightforward in supervised learning.

3.2.2 Disadvantages of Wrapper Method

Its drawbacks are:

- These methods have a higher risk of overfitting than filter techniques.
- Wrapper methods are computationally intensive.

Examples of this approach: Sequential forward selection, sequential backward elimination, beam search, genetic algorithms.

3.3 Embedded Method

Another type of feature selection technique is called embedded method. In this method the process of searching an optimal subset of features is included within the classifier construction, and is viewed as a search in the combined space of feature subsets and hypotheses. Embedded methods [25] are specific to a given learning algorithm like the wrapper methods. Advantage of these methods is that they include the interaction with the classification model like the wrapper methods and are less computationally intensive than wrapper methods. Examples of this approach are decision trees and artificial neural networks.

Examples of this approach: Decision trees, weighted NaiveBayes, feature selection using the weight vector of SVM. Table 1 shows the comparison of various feature selection methods [26].

Feature Selection Methods	Advantages	Disadvantages
Univariate Filter	Classifier independence, scalability and fast speed.	Lack of feature dependencies and classifier interaction
Multivariate Filter	Includes feature dependencies. Classifier independence. Better computational complexity.	Lack of classifier interaction, less scalable and slower.
Deterministic Wrapper	Includes classifier interaction, presence of feature dependencies	Classifier dependence, risk of overfitting.
Randomized Wrapper	Includes classifier interaction, presence of feature dependencies	Classifier dependence, high risk of overfitting, computationally intensive.
Embedded	Better computational complexity, includes classifier interaction	Classifier dependence.

Table 1: Comparison of Feature Selection methods

4. STEPS USED IN FEATURE SELECTION

This section discusses the steps followed in selecting the subset of features using filter approach. The steps used are:

- Initialize the learner.
- Load the dataset.
- Create the classifiers by training the learner on the dataset.
- Compute the relevance of the features.
- Set some margin, say m , and remove all those features for which relevance is less than m .

- Only the features whose relevance is greater than m are used for classification.
- Finally, use the learner on both the datasets and compare the accuracy.

5. EFFECTS OF FEATURE SELECTION ON VARIOUS DATASETS

The above mentioned steps were implemented on various machine learning algorithms (RandomForest, NaiveBayes and kNN) using different datasets used in our experiments. The experiments were carried out in Python programming using python machine learning tool. The datasets that were used in the experiments are the Australian Credit Approval dataset [19] from UCI Repository of Machine Learning Databases and Domain theories, Congressional Voting Records Dataset [20], and Adult Dataset [21]. The Credit dataset has already been used in the evaluation of the machine learning techniques earlier in our work [22].

5.1 Experiment

Different experiments were performed using different machine learners on the three datasets mentioned above. The machine learning algorithms used in the experiments are RandomForest, NaiveBayes and kNN. First the performance of the learning algorithms without feature selection has been shown on all the three datasets. After that performance of the learners has been checked by applying feature selection at margins 0.01, 0.02, 0.03 and 0.04. At different margins, it was observed that a number of irrelevant attributes got discarded depending on the calculated relevance of the attributes and only the relevant ones were used in the learning process.

Figure 1 shows the performance of machine learning algorithms on the Adult Dataset without using feature selection. Figure 2 shows the performance of machine learning algorithms on the same dataset using feature selection at margin 0.01. Figure 3 shows the performance of machine learning algorithms on the same dataset using feature selection at margin 0.03. Figure 4 shows the performance of machine learning algorithms on the same dataset using feature selection at margin 0.04.

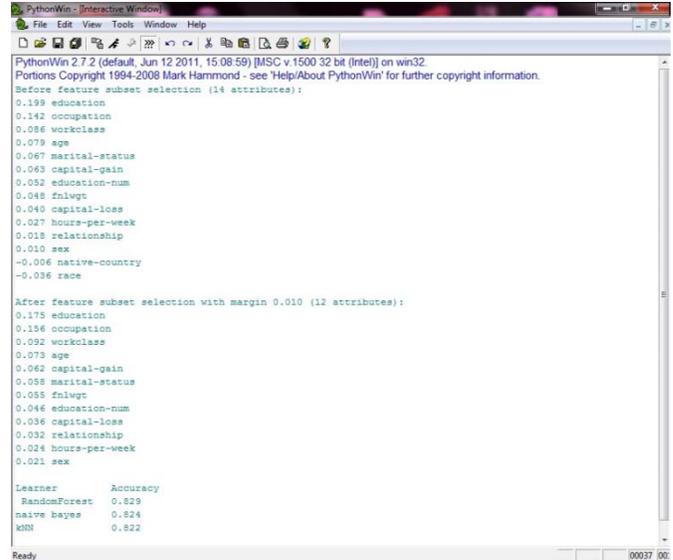


Figure 2: Results of Adult Dataset at margin 0.01

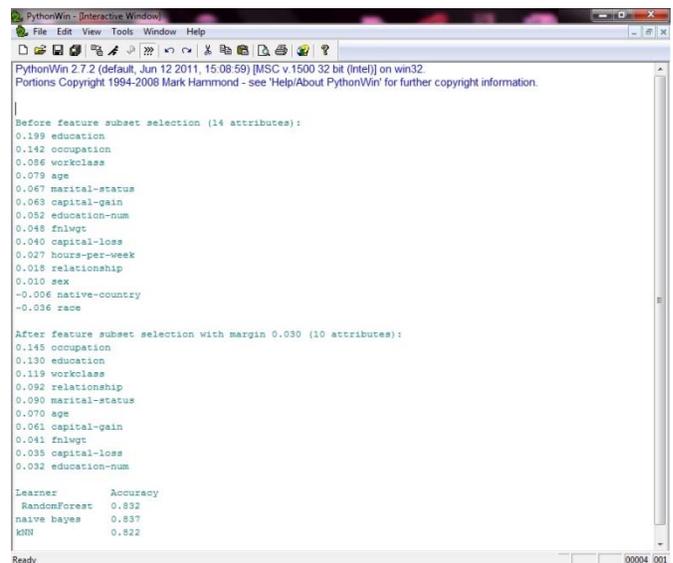


Figure 3: Results of Adult Dataset at margin 0.03

Table 2 shows the performance of the machine learners on the Adults Dataset in terms of accuracy, first without feature selection (FS) and then with feature selection at different margins. Table 3 shows the performance of the machine learners on the Credit Dataset in terms of accuracy, first without feature selection and then with feature selection at different margins. Table 4 shows the performance of the machine learners on the Voting Dataset in the same way.



Figure 1: Results of Adult Dataset without feature selection

```

PythonWin - Interactive Window
File Edit View Tools Window Help
PythonWin 2.7.2 (default, Jun 12 2011, 15:08:59) [MSC v.1500 32 bit (Intel)] on win32
Portions Copyright 1994-2008 Mark Hammond - see 'Help/About PythonWin' for further copyright information.

Before feature subset selection (14 attributes):
0.199 education
0.142 occupation
0.086 workclass
0.079 age
0.067 marital-status
0.063 capital-gain
0.052 education-num
0.048 fnlwgt
0.040 capital-loss
0.027 hours-per-week
0.018 relationship
0.010 sex
-0.006 native-country
-0.036 race

After feature subset selection with margin 0.040 (8 attributes):
0.182 education
0.161 occupation
0.093 workclass
0.074 relationship
0.071 marital-status
0.062 age
0.061 capital-gain
0.042 fnlwgt

Learner      Accuracy
RandomForest 0.803
naive bayes  0.818
kNN          0.830
    
```

Figure 4: Results of Adult Dataset at margin 0.04

Learners	Accuracy				
	Without FS	With FS (0.01)	With FS (0.02)	With FS (0.03)	With FS (0.04)
RandomForest	0.829	0.829	0.829	0.832	0.803
NaiveBayes	0.813	0.824	0.829	0.837	0.818
kNN	0.820	0.822	0.822	0.822	0.830

Table 2: Results of Adult Dataset at different margins

Learners	Accuracy				
	Without FS	With FS (0.01)	With FS (0.02)	With FS (0.03)	With FS (0.04)
RandomForest	0.845	0.852	0.838	0.837	0.850
NaiveBayes	0.864	0.864	0.858	0.851	0.830
kNN	0.831	0.835	0.831	0.851	0.831

Table 3: Results of Credit Dataset at different margins

Learners	Accuracy				
	Without FS	With FS (0.01)	With FS (0.02)	With FS (0.03)	With FS (0.04)
RandomForest	0.956	0.959	0.959	0.954	0.954
NaiveBayes	0.903	0.915	0.915	0.915	0.915
kNN	0.936	0.936	0.929	0.936	0.936

Table 4: Results of Voting Dataset at different margins

5.2 Results and Discussions

For all the datasets the learning algorithms have shown an increase in accuracy after feature selection. In some cases, the efficiency of learning algorithms after feature selection remained same as it was before feature selection depicting the fact that the discarded features were irrelevant and contributed nothing towards their performance as there was no change in accuracy even after discarding them. And hence were not needed. However, in some cases the learning algorithms show an increase in efficiency after feature selection. However, it increases only up to a certain limit. After that accuracy starts to decrease if feature selection is continued as more and more features are being discarded. Figure 1 and Table 2 show the results for Adult Dataset. Initially when no feature selection process is carried out, all the attributes of the Adult dataset (i.e. 14 attributes) are used in the learning process and the efficiencies of learners are 0.829 for RandomForest, 0.813 for NaiveBayes and 0.820 for kNN. After that when feature selection is carried out at margin 0.01, two of its attributes are discarded as their relevance is below the margin and thus only 12 attributes are used in the learning process. At this margin there is an increase in the accuracy of the learners e.g. NaiveBayes- 0.824 and kNN- 0.822 or it remains constant e.g. RandomForest- 0.829. However, at margin 0.04, four of its attributes are discarded and only eight attributes are used in the learning process. At this stage the accuracy starts decreasing. This shows that we have to find an optimal subset of features for a dataset. Similar effects have been shown with other two datasets as well wherein feature selection shows an increase in the accuracy of all the three machine learners up to a certain limit after which accuracy starts decreasing. Table 5 shows the number of attributes of all the datasets before feature selection (i.e. original number) and the number of attributes that were used in the learning process after feature selection.

Datasets	Number of Attributes at different margins				
	Before FS	0.01	0.02	0.03	0.04
Adult	14	12	12	10	8
Credit	15	11	6	5	4
Voting	16	14	14	13	13

Table 5: Number of attributes at different margins

The results in Table 5 show that in case of Adult dataset only 10 attributes among 14 were relevant because the learners showed better or similar performance even after discarding these 4 attributes. However, using only 8 attributes decreases the accuracy. In case of Credit dataset 11 out of 15 attributes were relevant and in case of voting dataset only 13 among 16 were relevant as it showed better or similar performance after discarding 3 of its attributes.

CONCLUSION

Different machine learning algorithms- NaiveBayes, RandomForest and kNN were used for evaluation on real data sets before and after feature selection taking into consideration user defined limits or margins (i.e. 0.01, 0.02, 0.03 and 0.04). The results have shown different effects on the accuracy rates while selecting the features at different margins. As we realized from the experiment that after increasing the margin beyond certain limit the performance starts degrading. Hence it is necessary to find an optimal subset of features for each dataset.

REFERENCES

- [1]. Zhang S., Zhang C., Yang Q. (2002), "Data Preparation for Data Mining", *Applied Artificial Intelligence*, Vol. 17, pp.375-381.
- [2]. Batista G. and Monard M.C. (2003), "An Analysis of Four Missing Data Treatment Methods for Supervised Learning", *Applied Artificial Intelligence*, Vol. 17, pp.519-533.
- [3]. Hodge V. and Austin J. (2004), "A Survey of Outlier Detection Methodologies", *Artificial Intelligence Review*, Vol. 22, Issue 2, pp. 85-126.
- [4]. Kira K. and Rendell L. (1992), "The Feature Selection Problem: Traditional Method and a New Algorithm", in *Proceedings of the Tenth National Conference on Artificial Intelligence*, pp. 129-134. MIT Press,
- [5]. Liu H. and Motoda H. (1998), "Feature Extraction, Construction and Selection: A Data Mining Perspective". *Kluwer Academic Publishers*, Boston/Dordrecht/London, Boston.
- [6]. Kohavi R. and John G. (1997), "Wrappers for Feature Subset Selection". *Artificial Intelligence'97*, pp. 273-324.
- [7]. Almuallim H. and Dietterich T. (1991), "Learning with Many Irrelevant Features", *Proceedings of the Ninth National Conference on Artificial Intelligence*, pp. 547-552. MIT Press.
- [8]. Quinlan J. (1993), "C4.5: Programs for Machine Learning", *Morgan Kaufmann*, San Mateo.
- [9]. Moore W. and Lee S. (1994), "Efficient Algorithms for Minimizing Cross Validation Error", *Machine Learning: Proceedings of the Eleventh International Conference*.
- [10]. John G., Kohavi R. and Pfleger K. (1994), "Irrelevant Features and Subset Selection Problem", *Proceedings of the Eleventh International Conference on Machine Learning*, pp. 121- 129. Morgan Kaufmann Publishers.
- [11]. Caruana R. and Freitag D. (1994), "Greedy Attribute Selection", *Machine Learning: Proceedings of the Eleventh International Conference*, W. Cohen and H. Hirsh (eds). Morgan Kaufmann.
- [12]. Weston J., Mukherjee S., Chapelle O., Pontil M., Poggio T. and Vapnik V. (2001), "Feature selection for SVMs", *Advances in Neural Information Processing Systems 13*. Cambridge, MA: The MIT Press, pp. 668-674.
- [13]. Forman G. (2003), "An extensive empirical study of feature selection metrics for text classification", *Journal of Machine Learning Research*, Vol. 3, pp. 1289-1305.
- [14]. Guyon I. and Andr'e E. (2003), "An introduction to variable and feature selection", *Journal of Machine Learning Research*, Vol. 3, pp. 1157-1182.
- [15]. Dash M. and Liu H. (1997), "Feature selection for classification", *Intelligent Data Analysis*, Vol. 1(1-4), pp. 131-156.
- [16]. Yang Y. and Jan O. P. (1997), "A comparative study of feature selection in text categorization", *ICML '97: Proceedings of the Fourteenth International Conference on Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 412-420.
- [17]. Talavera, L. (2005), "An evaluation of filter and wrapper methods for feature selection in categorical clustering", *Advances in Intelligent Data Analysis VI Springer-Verlag*, pp. 440-451.
- [18]. Saeys Y., Inza I. and Larranaga P. (2007), "A review of feature selection techniques in bioinformatics", *Bioinformatics* Vol. 23 No. 19, pp. 2507-2517.
- [19]. Australian Credit Approval dataset <http://www.hakank.org/weka/credit.arff> from UCI Repository of Machine Learning Databases and Domain theories (<http://archive.ics.uci.edu/ml/datasets.html>). Accessed on 20-04-2013.
- [20]. Congressional Voting Records Dataset (<http://archive.ics.uci.edu/ml/datasets/Congressional+Voting+Records>). Accessed on 20-04-2013.
- [21]. Adult Dataset <http://archive.ics.uci.edu/ml/datasets/Adult>. Accessed on 20-04-2013.
- [22]. Khan M. and Quadri S.M.K. (2012), "Evaluating Various Learning Techniques For Efficiency", *International Journal of Engineering and Advanced Technology (IJEAT)*, Vol. 2, Issue 2, pp. 326-331.
- [23]. Zenglin X., Rong J., Jieping Y., Michael R. L. and Irwin K. (2009), "Discriminative semi-supervised feature selection via manifold regularization", *IJCAI' 09: Proceedings of the 21th International Joint Conference on Artificial Intelligence*.
- [24]. Brown G., Pocock A., Zhao M.J. and Lujan M. (2012). "Conditional Likelihood Maximisation: A Unifying Framework for Information Theoretic Feature Selection", *Journal of Machine Learning Research (JMLR)*.
- [25]. Canedo V.B., Maroño N.S. and Betanzos A.A. (2013), "A review of feature selection methods on synthetic data", *Springer-Verlag, Knowl Inf Syst* Vol. 34, pp. 483-519.
- [26]. Ladha L. and Deepa T. (2011), "Feature Selection Methods and Algorithms", *International Journal on Computer Science and Engineering (IJCSSE)* Vol. 3 No. 5, pp. 1787-1797.

- [27]. Yang P., Liu W., Zhou B.B., Chawla S. and Zomaya A.Y. (2013), “Ensemble-Based Wrapper Methods for Feature Selection and Class Imbalance Learning”, *Springer-Verlag* pp. 544-555.
- [28]. Auffarth B., Lopez M. and Cerquides J. (2010), “Comparison of Redundancy and Relevance Measures for Feature Selection in Tissue Classification of CT images”, *Advances in Data Mining Applications and Theoretical Aspects*, *Springer* pp. 248-262.

Path Optimization Using APSO

Deepak Goyal¹ and Malay Ranjan Tripathy²

Submitted in March, 2013; Accepted in August, 2013

Abstract - This paper addresses the malicious node detection and path optimization problem for wireless sensor networks. Malicious node detection in neighborhood is a needed because that node may cause incorrect decisions or energy depletion. In this paper APSO (combination of Artificial bee colony and particular swarm optimization) is used to choose an optimized path. Through this improved version we will overcome the disadvantage of local optimal which comes when we use PSO approach.

Index Terms - Green IT, Environmentally Sustainable, Environmental Intelligence (EI), Server Virtualization, Cloud Computing

1. INTRODUCTION

Wireless networks or sensor networks are composed of a large no. of dynamically located sensor nodes. The sensors collect the data and forward to the base station through defined communication path. Data is forwarded from source to sink node. If the information is sensitive, the nodes and communication path must be trust worthy. If any node in path is suspicious node need to calculate the alternative path.

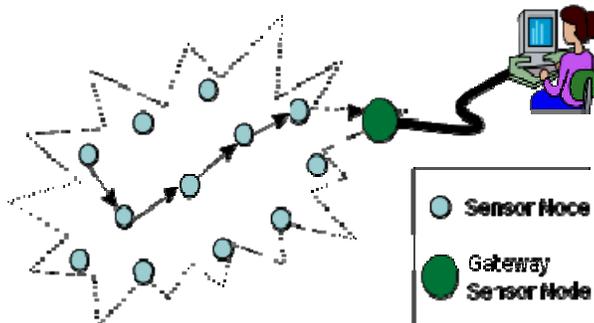


Figure 1: Wireless Sensor Network

Node may be suspicious due to internal reason like traffic load or external factors like temperature. There are many methods to calculate the trust of a node like Reputation-based, Event-based, Agent-based, etc.

As shown in figure 2 there are multiple paths from source to sink node but we have to choose optimized path. Now after detection of malicious node an alternate path is needed. There are so many strategies to find alternative path. Swarm Intelligence is the one of them. Swarm Intelligence itself has categories like Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Artificial Bee Optimization (ABC), etc.

Routing protocols may maintain single or multiple routes to a destination node. Single path protocols can find one or Multiple routes and so select the best path for data transport, discarding other ones and multipath routing refers to the protocols that find, maintain, and use those paths to transport sensed data [1].

Classical based Routing and Swarm Intelligence are one of the fields for path optimization. In [2] Authors compared many of these protocols. In this paper we would consider SI field.

2. RELATED WORK

2.1 PSO

Particle swarm optimization (PSO) is a popular multidimensional optimization technique. Ease of implementation, high quality of solutions, computational efficiency and speed of convergence are strengths of PSO. Advantages of PSO:

- 1) Ease of implementation on hardware or software.
- 2) Availability of guidelines for choosing its parameters.
- 3) Availability of variants for real, integer and binary domains.
- 4) Quick convergence [3].

In [4] paper, particle swarm algorithm was used to find the optimal positions of the sensors to determine the best coverage.

In [5] the modified form of PSO by the usage of an explicit consensus mechanism is used for optimization.

In [6] Authors consider the maximization of the coverage as an optimization criterion by implementing a centralized technique. Their technique is based on a modified PSO strategy they called their technique Particle Swarm Genetic Optimization (PSGO).

In [15] author considers Sensor Deployment Problem using Particle Swarm Optimization (PSO). This work has the ability to achieve optimal solution of coverage problem with minimum number of sensors in wireless sensor networks. This approach cultivates an innovative idea in employing the PSO algorithm with enhanced fidelity. The results show that the PSO approach is effective and robust for efficient coverage problem of sensor deployment and is considered to give almost the optimal solution in WSN.

In [16] the modified form of PSO by the usage of an explicit consensus mechanism is used for optimization. Author said that it is not useful to consider a global best position, because it implies a centralized scheme of control or, at least, the capacity of the nodes to communicate with every other node in the sensor field. In order to take into account the limited communication capabilities of sensors, we stated that the social term involves the position that enjoys the maximum consensus within each node's neighborhood, where a neighborhood is composed only of the sensors within its transmitting/receiving range.

In [17] author proposed a virtual force co evolutionary PSO for dynamic deployment of Nodes. Virtual force based dynamic

¹Jagannath University, Jaipur, ²Amity University, Noida
E-mail: deepakgoyal.vce@gmail.com and amity@gmail.com

deployment involves iteratively moving a sensor based on virtual attractive or repulsive forces from other nodes, obstacles in the field and the areas that need higher coverage probability. Virtual force vectors depend on the distance between nodes and whatever attract or repulse them, and their relative directions. A sensor's new positions are computed in such a way that it moves in the direction of the virtual force by a step size proportional to its magnitude.

In [18] author used multi-base for optimal positioning of base station in a two tier WSN [17]. The two tier network consists of nodes that can communicate only with the application nodes they are assigned to. Application nodes possess long-range transmitters, high-speed processors, and abundant energy. The PSO Multi-Base method aims at determining positions of base stations so that the total of distances of application nodes to their nearest base stations is minimum.

In [19] Authors consider the maximization of the coverage as an optimization criterion by implementing a centralized technique. Their technique is based on a modified PSO strategy they called their technique Particle Swarm Genetic Optimization (PSGO).

2.1.1 Disadvantage of PSO

But there is a disadvantage of PSO which is local optimal because particle get flung away from the best location since global best value may be updated above a certain value so we need to keep track of how many iteration has passed since global best is updated.

2.2 ABC

It simulates the artificial bees to find out the best nectar source with swarm intelligence.

In [7] author applied the ABC algorithm to the dynamic deployment problem in WSNs with mobile sensors.

In [10] Author models the behavior of social insects, such as ants, birds or bees, for the purpose of search and in it problem solving has been the emerging area of swarm intelligence. Honey-bees is most closely studied social insects. Here, an artificial bee colony algorithm is developed to solve clustering problems which is inspired by the bees' forage behavior.

3. PROPOSED APPROACH

3.1 Detection of Malicious Node

If any node in the path is suspicious, that node is malicious and it is need to be detected Wireless sensor networks or sensor networks are composed of a large number of sensor nodes deployed densely in a closed proximity to collect data to a specific function. There are varieties of methods to calculate the trust of a successive node. The methods include the reputation-based trust management, event-based trust management, collaborative trust management, and agent-based trust management [8].

We would use Agent Based approach and Event Based approach.

3.2 Agent-Based Approach and Event-Based Approach:

Agent-based trust model for WSN can't distinguish different events which effect trust rating, and all the events have the same affects. Here, we propose a method in which different events considers to detect the nodes which can be faulty in different events.

In our trust framework, the trust rating is dependent on different events of the sensor nodes in WSN. It means that at different event, the node has different trust rating, which also means a sensor node has several trusting rating stored in its neighbor nodes.

Our trust framework runs at the agent node which has strong competence to compute, large storage and memory. The agent node uses Threshold value to monitor all kind of event happened in sensor nodes within its radio range and functions in a completely distributed manner. Threshold value is taken to be 4, if 4 or more packets comes to agent node then it would be considered as the high traffic and node can become faulty which would be protected by choosing another path as described in next section. Every agent node maintains trust table for nodes. In our framework, a node has several trusts rating value which would be stored at agent node. The number of trust rating in a sensor node depended on the number of events in sensor node. Here we store event trust of a node in a binary format. Here considers positive event and 0 shows negative event. This high traffic detection is needed because every node has some defined energy and node use this energy to send packets.

If heavy traffic attack to an agent node then energy of an agent node decreases suddenly and node may becomes faulty or maliciousness. In our work we set threshold such that we can stop traffic to an agent node before it becomes faulty.

In our trust framework first only one node tries to send packet to agent node as shown in figure 2. That event is trusted.

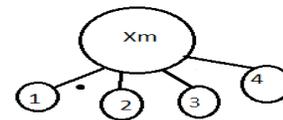


Figure 2: Event 1 where only one packet enters the agent node

Now let at Event 2, 4 packets tries to enter the agent node as in figure 4shown and we have set the threshold value 4 so now to save it from becoming dead node we would change the paths of all nodes which were used to go through it because now this node is at its peak value it will be dead as one more node will come to this node.

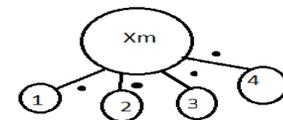


Figure 3: Event 2 where 4 Packets try to enter the agent node

In this paper after malicious node detection that node will inform its neighbors of its maliciousness and suggest another global best to calculate alternate path using modified version of PSO which

would be called APSO (Artificial Bee Particle Swarm Optimization). Addition of ABC will cover the limitation of PSO.

3.3 Particle Swarm Optimization

PSO algorithm works by having a population of particles.

Let N is the no. of particles in swarm, each having a position x_i in the search-space and a velocity v_i . Let p_i be the best position of the particle i and let g_b be the best position of the whole swarm.

Algorithm is:

For each particle $i = 1$ to N :

Initialize particle's position with uniformly distributed random vector: $x_i \sim U(b_{lo}, b_{up})$, b_{lo} and b_{up} are the lower and upper boundaries of the search-space. Initialize the particle's best position to its initial position: $p_i \leftarrow x_i$.

If $(f(p_i) < f(g_b))$ update the swarm's best position: $g_b \leftarrow p_i$

Initialize particle's velocity: $v_i \sim U(-|b_{up} - b_{lo}|, |b_{up} - b_{lo}|)$ Until an end criterion is met, repeat:

Step1.Count = 0

For every particle $i = 1$ to N :

For every dimension $d = 1$ to n :

Step2. Pick the random numbers: $r_p, r_g \sim U(0,1)$

Step3.Update particle's velocity: $v_{i,d} \leftarrow v_{i,d} + \varphi_p r_p (p_{i,d} - x_{i,d}) + \varphi_g r_g (g_{b,d} - x_{i,d})$

Update particle's position: $x_i \leftarrow x_i + v_i$

Step4. If $(f(x_i) < f(p_i))$:

Update particle's best position: $p_i \leftarrow x_i$

Step5.If $(f(p_i) < f(g_b))$ update swarm's best position: $g_b \leftarrow p_i$

Step6.Now g_b holds the best solution. And call it Ta_best .

Count = count+1.

The parameters φ_p , and φ_g are to be selected.

The function used here is sphere function whose global minimum value is 0 at $(0, 0, \dots, 0)$. It is a unimodal function with non-separable variables.

$$F(x) = \sum x^2$$

3.3 APSO

The Process of APSO is as:

Step 1 Initialization of Parameters: set number of individuals of the swarm; set maximum circle-index; set other constants needed.

Step 2 Initialization of the colony: firstly, generate a colony with specific number of individuals. Then as a bee colony, it is divided into two categories, each individual's fitness value; on the other hand, as a particle swarm, calculate the fitness value of each particle and take the best location as the global best location. We assume that the cyclic number is represented by $iter$, and $iter + 1$.

Step 3 Perform Particle Swarm Optimization. The best location in the iteration will be called Ta_best . There is a count variable which will be updated for each updation of the global best.

Step 4 If the global best is greater than 2 then run the Artificial Bee Colony Algorithm. After all the choices above have been made, the best solution is generated in this iteration which we called it GlobalMin.

Step 5 The minimum between the value of GlobalMin and the value Ta_best is GlobalMins and is defined by the following equation:

$$\begin{aligned} \text{Globalmins} &= \text{globalmins}, \text{ if } \text{globalmins} \leq \text{Ta_best} \\ \text{globalmins} &= \text{Ta_best}, \text{ if } \text{Ta_best} \leq \text{globalmins} \end{aligned}$$

And the GlobalMin and the Ta_best will both be equal to the value GlobalMins, and will be substituted into next iteration $iter = iter + 1$.

Step 6 If the number of circles is greater than the maximum of circle - index. If not, go to Step 2; if it is, end the process and save the value GlobalMins.

3.4 Artificial Bee Colony

It simulates the artificial bees to find out the best nectar source with swarm intelligence. Just like the artificial bee colony in reality, at this algorithm, all the artificial bees are mainly divided into two categories. One is called employed foragers. Their job is to gather honey from their corresponding nectar source, and to exchange information of their source with other bees. The specific employed foragers whose source is the best at the present will become the ones who lead others to their source. The other one is the unemployed foragers. They are the one who don't find out the suitable source by themselves. They can keep looking for the source or follow the lead foragers to gather honey. There are scouts, search the surrounding the nest and they try to find new food sources. The source is suitable or not is decided by the fitness value. The larger the fitness value, the better the sources is.

4. RESULTS

4.1 Parameters

Energy: As each node in the network is a sensor node, each node is defined with specific energy we have defined 6 Joules to each node. With each communication over the network some energy is lost. If the energy is less than minimum required energy or 0, the node will be dead itself. Here we keep Threshold 4 so as to prevent node from becoming dead.

Number of Packets: This property represents the number of successful packet coming to a cluster head for a specific communication.

Pbest: Particles own experience. It is location which particles remember where it was closer in the past means Particle knows its own best position.

Gbest: Whole swarm's best. The movements of the particles are guided by their own best known position in the search-space as well as the entire swarm's best known position. Particle choose minimum of the two.

Ta_best : Minimum of Pbest and Gbest.

Velocity: Initial velocity of packet is taken to be 5m/s. Particle changes its position by updating its location and velocity.

$\varphi_p : 1/3$

$\varphi_g : 1/2$

Globalminimum: Employed foregoers, they are keeping searching the new sources around them and determine which one

is turned to according to the minimum fitness value of each source which we called it GlobalMin.

Threshold: We have taken 4 as its value at the time of detecting malicious node. When this value comes node is said to be malicious.

Count: We have taken its value to be 2. It is used to check how many times global best is updated. When this value exceeds colony is updated from Particle Swarm Optimization to Artificial Bee Colony so that particle does not trap in local optimal problem.

Count1: Shows the no. of times a packet can go to the sink nodes using PSO in the same time as compared to APSO.

Count2: Shows the no. of times a packet can go to the sink nodes using APSO in the same time as compared to PSO.

4.2 Results and Discussion

A node wants to send a packet to its sink node. For it node send packet to its agent node which would send further agent of sink node (figure 6). This is called event 1 so it is a Positive event.

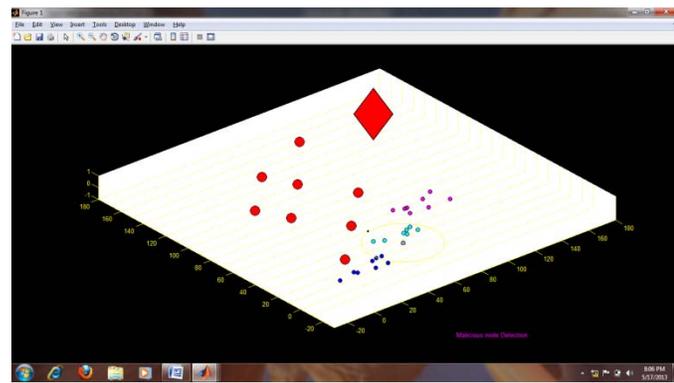


Figure 4

As discussed in earlier section that this trust framework runs at the agent node which has strong competence to compute, large storage and memory. The agent node uses Threshold value to monitor all kind of event happened in sensor nodes within its radio range and functions in a completely distributed manner. Threshold value is taken to be 4, if 4 or more packets comes to agent node then it would be considered as the high traffic and node can become faulty which would be protected by choosing another path as described in next section. Every agent node maintains trust table for nodes. So in this paper we use high traffic load as a malicious measure. If packets greater than threshold value which is 4 try to enter the agent node to send their packets to sink node. So it cause high traffic load on agent node (figure 7). This is event 2 and is negative event.

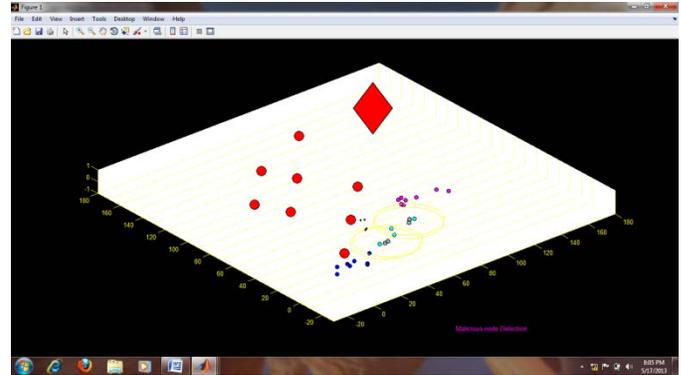


Figure 5

Now Event – Based approach is applied to the agent node and node becomes yellow to show high traffic load. As energy gets lost with each communication agent node would not be able to tackle all packets at a time because this may cause fault to the agent node and so if value of threshold reaches agent node stop working.

The second agent node is detected as bad nodes (shown in yellow). This bad node does not allow pass packets to the next nodes. It informs its neighbour nodes to choose another path to sink node. And by applying APSO we calculated new path. So, aafter the bad node is detected, packets changed their path to reach the destination based on APSO. Firstly our algorithm use concept of PSO (figure 8).

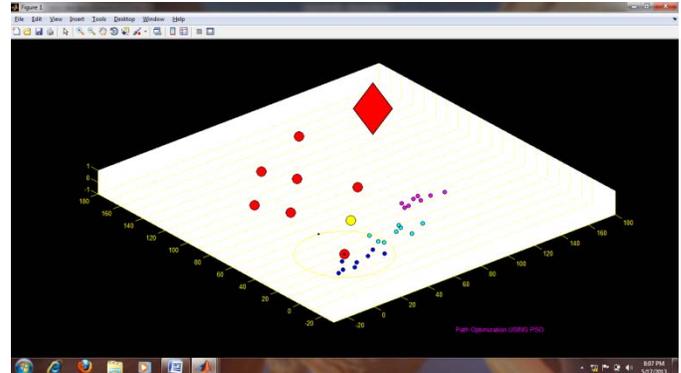


Figure 6

As we have used count limit is 2 so after it becomes two it uses the concept of Artificial Bee as there is possibilities of coming more new nodes which may lead to better path. As in Artificial bee scout helps in finding new source. Here both PSO and ABC find their best and these are compared and we choose the best one for next location. And now local optimum (discussed in earlier section) of PSO is removed (figure 9). New nodes are represented by cyan colour.

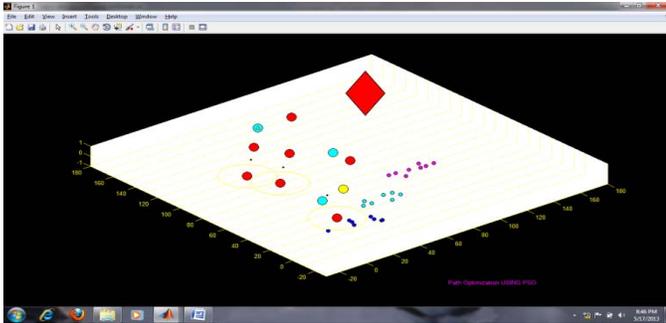


Figure 7

Now Artificial bee found new better route and now it can be used (figure 10).

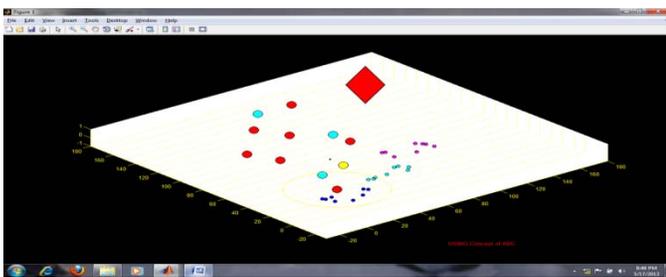


Figure 8

Now we calculate the no. of times a packet can go to the sink nodes using APSO in the same time as compared to PSO. We would count it through two variables count1 (for PSO), count2 (for APSO). This is shown in table1 in next part below.

4.3 Graph of Malicious node Detection

At event 1 as discussed in results it is positive event and we have chosen 1 as a trust value for positive event.

At event 2 as discussed in result it is negative event and it is not trusted one and we will show it as 0 value for trust value of node which will force us to change the direction to sink node and we have chosen APSO algorithm to do that. This is shown in figure 11.

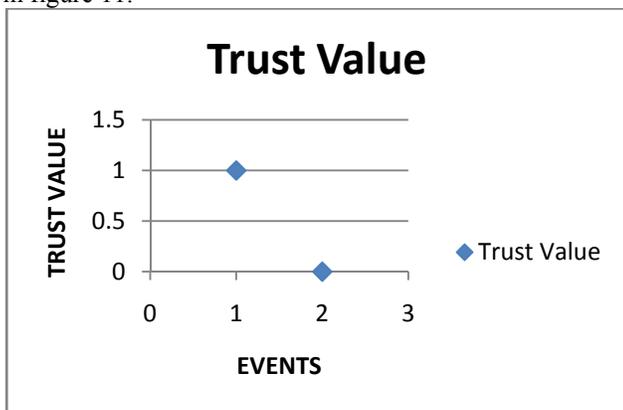


Figure 9: Trust Rating

4.3.1 Comparison of PSO and APSO

As we have discussed before there is a limitation of PSO that it can trap in local optimum. If after trapping in local optimum of PSO we use only PSO the following graph is made. Figure 12 shows the distance travelled by packets.

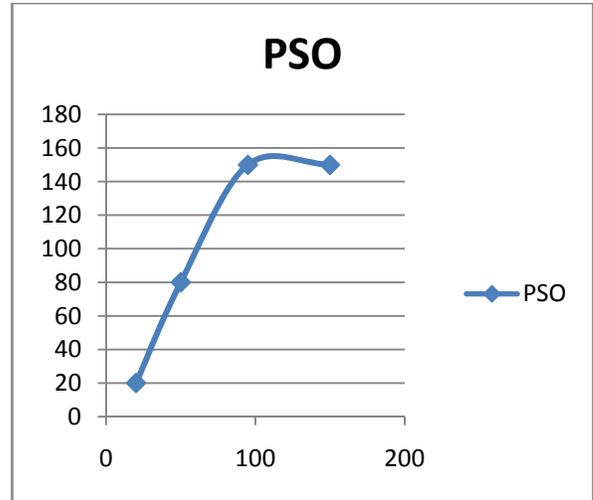


Figure 10: Path chosen by PSO

But if we use our APSO we found that a new better path has come and this makes packet to go through a best shortest path as shown in following graph. Figure 13 shows distance travelled by packets. So now we can say that our proposed Algorithm (APSO) is better than PSO.

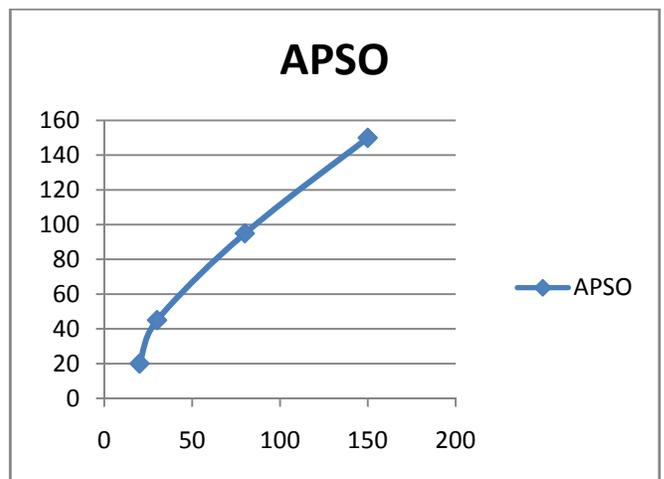


Figure 11: Path chosen by APSO

Chosen Strategy	Packet Delay	Value of count1, count2
PSO	More (because packets are going through long path and APSO can	Count1=1

	send two times in the same time of PSO)	
APSO	Less	Count2=2

Table 1: Comparison of packets delay in PSO and APSO

After implementing APSO packets transmitted (figure 14) over the network is increased as compared to PSO (figure 15) in minimum time because packets are transmitted through the shortest path. Packet delay is shown in table 1. This is counted by count1 (for PSO) and count2 (for APSO) variables in our proposed work.

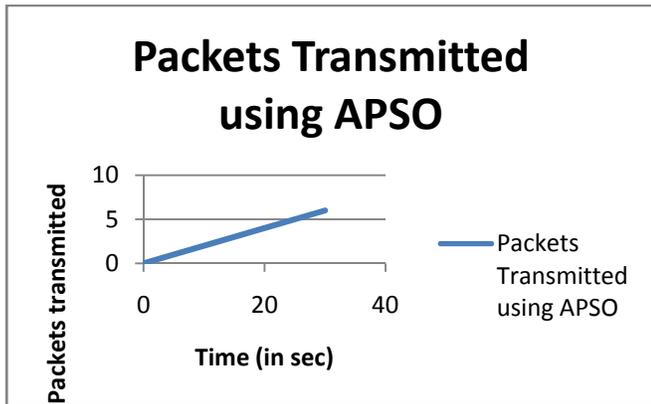


Figure 12

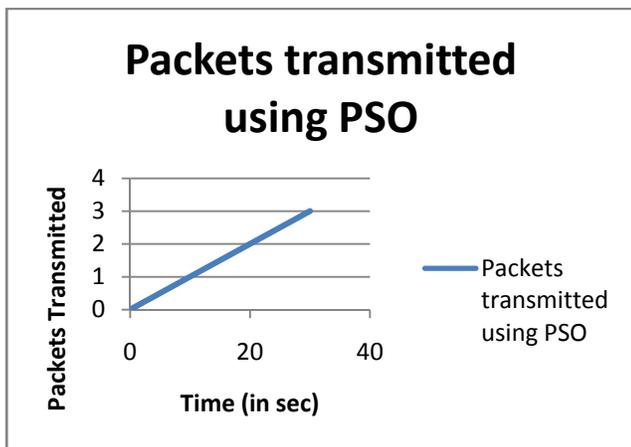


Figure 13

5. CONCLUSION

In this paper we consider high traffic load as a measure of malicious node. It is nature of these networks that higher traffic load is observed in some events. Here we use Agent-based approach with different events for detection of node before it becomes a faulty node and that detected node informs its entire neighbour about its maliciousness to choose another path and also we use an combine approach of Artificial bee colony and PSO called APSO to choose on optimized path. After detection of malicious node due to high traffic load we calculated

alternate path using modified PSO (APSO). As a result APSO can send two times more packets in the same times of PSO, we use our APSO we found that a new better path has come and this makes packet to go through a best shortest path and thus it overcomes the disadvantage of PSO which is used to get trap in the local optimal.

REFERENCES

- [1] Gianni A. Di Caro, Muddassar Farooq and Muhammad Saleem, "Swarm Intelligence Based Routing protocol for Wireless Sensor Networks: Survey and Future Directions", ELSEVIER, Information Sciences, Volume 181, Issue 20, 15 October 2011, pp. 4597-4624.
- [2] Adamu MurtalaZungeru, KahPhooiSeng and Li-MinnAng, "Classical and swarm intelligence based routing protocols for wireless sensor networks: A survey and comparison", Journal of Network and Computer Applications 35, 2012, 1508-1536.
- [3] Ganesh Kumar Venayagamoorthy and Raghavendra V. Kulkarni, "Particle Swarm Optimization in Wireless Sensor Networks: A Brief Survey", IEEE Transactionson Systems, MAN, and CYBERNETICS, March 2010.
- [4] Babu Rao Thella, Nikitha Kukuluru, and Rajya Lakshmi Davuluri, "Sensor Deployment Using Particle Swarm Optimization", Nikitha et. al. / International Journal of Engineering Science and Technology, vol. 2(10), 2010.
- [5] Enrico Natalizio, Francesca Guerriero, Valeria Loscri, "Particle Swarm Optimization Schemes Based on Consensus for Wireless Sensor Networks", MSWiM, 2012.
- [6] J. Li, K. Li, and W. Zhu, "Improving sensing coverage of wireless sensor networks by employing mobile robots", in Proc. Int. Conf. Robot. Biomimetics, pp. 899-903, 2007.
- [7] Beyza Gorkemli, Celal Ozturk and Dervis Karaboga, "Artificial Bee Colony Algorithm for Dynamic Deployment of Wireless Sensor Networks", Turk J Elec Eng & Comp Sci, Vol.20, No.2, 2012.
- [8] Yenumula B. Reddy, "Trust-Based Approach in Wireless Sensor Networks Using An Agent to Each Cluster", International Journal of Security, Privacy and Trust Management (IJSPTM), Vol.1, No.1, February 2012.
- [9] Alex Doli, Constantin Volosencu, Daniel-Ioan Curiac, Octavian Dranga and Tomasz Bednarz, "Neural Network Based Approach for Malicious Node Detection in Wireless Sensor Networks", Proceedings of the 2007 WSEAS Int. Conference on Circuits, Systems, Signal and Telecommunications, Gold Coast, Australia, January 17-19, 2007.
- [10] Changsheng Zhang, Dantong Ouyang, Jiaxu Ning, "An artificial bee colony approach for clustering", ELSEVIER, Expert Systems with Applications 37, 2010, 4761-4767.

A Study of Empowerment and Engagement of ITES/BPO Professionals in Delhi-NCR

Puja Khatri¹ and Khushboo²

Submitted in April, 2013; Accepted in August, 2013

Abstract - With the advent of globalization, technology revolution has fastened its pace. Owing to availability of skilled, English speaking and cheap manpower, India enjoys being the favorite destination for outsourcing. These organizations are characterized by IT enabled operations, strategic HR practices, rotational shifts and high employee turnover rate. Despite good remuneration and other benefits, this sector witnesses maximum attrition. Managers try to empower and support their subordinates but the engagement level of employees is decreasing day by day. The present study seeks to understand the perception of employees pertaining to the level of empowerment and engagement with respect to variables like, job security, work-life balance, concern from top management, performance review system etc.. The study was conducted in Delhi NCR region wherein the respondents were selected through multistage sampling (N=100). Data was collected through self-constructed questionnaire (cronbach alpha =0.83). Demographic differences were also explored as male and female employees are treated differently at BPOs. Relationship between the level of empowerment and engagement was also analyzed. The paper seeks to provide an insight to top management and corporate practitioners of BPOs, who can use advanced techniques to develop an empowered culture thus, promoting engagement.

Index Terms – Attrition in BPOs, Empowerment of BPO employees, Engagement, Gender issues in BPOs, ITES.

1. INTRODUCTION

India is full of educated, English speaking and tech-savvy graduates who are ready to work in fewer salaries as compared to their western counterparts. Indian IT sector offers cost effective services to the customers situated globally and has emerged as one of the major key players in this global village. Companies readily assign their non core functions to the outsourcing firms in order to focus on their core competence. Such companies to which these functions are assigned are known as BPOs (Business Process Outsourcing) (Pathak and Sarin, 2011) [1]. Availability of skilled and relatively cheap manpower makes India a favorite destination for running operations of ITES/BPO sector. By the advent of technology revolution on global front, organizations are adopting ITES for almost every aspect of their employees make huge money out of it but on its contrary, it leads to high turnover rate as well

(Desai, Majumdar and Prabhu, 2010) [2]. Present workforce is of knowledge workers who want to upgrade their skills and knowledge on continuous basis. They want respect at workplace, fair treatment, compensation according to their contribution, transparent reward policies etc. In the rise of such events, the role of HR has evolved from that of personnel management to strategic HRM. HR managers are striving for betterment in this sector. They have initiated incentive schemes, socialization ceremonies, career development programs, reward and recognition programs, open door policy etc. Infrastructure of such companies is outstanding and workstations are ergonomically designed. Still this sector needs reinvention in terms of transparent practices, fair workload and reward distribution and competitive salary packages. Meaningful work is said to have great influence on the cognitive part of performance (Spreitzer, 1995) [3]. Also, role of transformational leadership is crucial. The way supervisors behave can mould the performance level of employees. Poaching is another issue faced by this sector (Mishra, 2007) [4]. This happens in absence of sufficient growth opportunities in the organization. Organizations need to understand that individual growth matters to employees. Govt. of India along with State Govts. are taking significant steps to promote India as the most preferred destination for ITES. NASSCOM is helping govt. in formulating certain policies and frameworks in order to revive this industry (Begum, 2013) [5].

This study is an attempt to understand the underlying factors affecting empowerment and engagement level of employees and to establish a relationship between two. The findings of this study will provide invaluable insights to BPO managers so that they can redesign or improve the work patterns.

2. LITERATURE REVIEW

Business Process Outsourcing is the branch of IT by which services are rendered to the companies' clients. These services can be customer care, handling IT operations, financial backend processes etc. In present context India's competence in IT has been widely recognized and appreciated in ITES sector majorly. In India, BPO industry has grown in the last decade. India enjoys being the favorite offshore destination because of people (manpower) and location. Over 2million students graduate every year in India who are computer literate and can speak English. They are willing to work at nearly 80% less salary than their western counterparts (Raman, Budhwar and Balasubramanian, 2007) [6]. BPOs have a formal and structured HRM system in place which performs many strategic functions to enhance involvement and commitment of employees (Raman et al, 2007) [6]. The extent to which an employee is involved in one's task will enhance the engagement level of employee with the

^{1, 2}USMS, Guru Gobind Singh Indraprastha University, Delhi, India

E-Mail: ¹pujakhatri12@gmail.com and

²khushboo0803@gmail.com

organization. Evolution of the concept of employee engagement can be traced back to 1990s when Kahn (1990) [7] explained its meaning as “*the harnessing of organization members’ selves to their work roles; in engagement, people employ and express themselves physically, cognitively, and emotionally during role performances* (p.694)”. The concept of engagement may be defined by different organizations differently but the end result is same for all and i.e. discretionary effort (Perrin, 2003) [8]. Devi (2009) [9] opines that engagement is the degree to which an employee puts discretionary efforts into his or her work over and above required time, brainpower or energy. Apart from competitive remuneration, teamwork, healthy working conditions, peer relationship, skill development, advancement opportunities and learning opportunities are said to have an impeccable effect on the level of engagement. Saks (2006) [10] conducted a study in order to test a model of the antecedents and consequences of job and organization engagements based on social exchange theory. He identified that engagement can be at two levels: job engagement and organization engagement and concluded that different factors act as their antecedents. Perceived organizational support predicts both; job characteristics act as antecedent to only job engagement and procedural justice predicts organization engagement. Job and organization engagement mediated the relationships between the antecedents and job satisfaction, organizational commitment, intentions to quit, and organizational citizenship behaviour.

Pritchard (2008) [11] opines that the concept of Employee Engagement can be fairly defined in terms of: Say (Measure of how employees are going to portray their own organization), Stay (measuring employees’ loyalty towards their organization) and Strive (measures employees’ readiness to do “over and above” their duties to attain organizational success). These three dimensions so mentioned are clearly relevant for the BPO workforce. May, Gilson and Harter (2004) [12] opine that meaningful jobs, job-security and presence of transformational leaders are significantly related to engagement. Maslach, Schaufeli and Leiter (2001) [13] argue that job engagement is associated with a fair allocation of workload, autonomy to exercise choice and control, receiving recognition and rewards for efforts, a supportive work environment, fairness and justice, and meaningful and valued work. IT managers largely depend on their subordinates for understanding the technical solutions to propose strategies for business. They need to play the role of a coach and communicator in order to have coordination (Kanooni, 2005) [14]. The support extended by the supervisor indeed enhances the engagement level of employees.

Studies have revealed that engaged employees tend to be more productive, efficient and have more chances to stay with their organization. Employee engagement has a positive correlation with enhanced performance. According to Lockwood (2007) [15], employee engagement acts as a strong antecedent to organizational success. Greater the engagement greater will be the employee retention, loyalty, customer satisfaction and enhanced organizational reputation (Lockwood, 2007; Gallup, 2008; Desai et al, 2010 and Hallberg and Schaufeli, 2006) [15]

[16] [2] [17]. Engaged employees are capable of propelling organization to new heights of growth. A proper communication channel in place, employee participation and empowerment are very crucial elements of employee engagement. An empowered employee base is considered to be more engaged resulting in improved productivity, motivation level and better aligned with the organizational objectives. Such employees have less or no intention to leave their organization. In a study conducted by Stander and Rothmann (2010) [18], it has been found that Psychological empowerment predicts Engagement. It emphasizes that employees who recognize a purpose in their work, recognize their own caliber and skills, capable of pursuing their career goals and believe that they can control the system in which they are working, are more engaged as compared to their peers. The dictionary meaning of “Empower” is to promote the self-actualization or influence of or to give official authority or legal power to. Spreitzer (1995) [3] defines empowerment as an individual’s experience of internal motivation that is based on cognitions about oneself in relation to one’s job role. He also identified four individual psychological determinants which may affect behavior of an employee viz. meaning, competence, self determination and impact. These four determinants might be considered as psychological conditions that lead to engagement (Stander et al, 2010) [18]. The final goal of empowerment is to create and empowered organizational culture, which is time consuming and demands dedication of top management and involvement of employees (Kanooni, 2005) [14]. One cannot deny the importance of leadership in creating an empowered culture. Leaders exhibiting empowerment behavior are capable of fostering success as employees feel empowered through enhanced responsibility, authority to take decisions, receiving timely information and feedback, motivation and encouragement. Elmuti (1997) [19] opines that empowerment process can’t be initiated where autocratic form of leadership exists. So, role of leadership can’t be ignored. BPO industry is struggling to fill the gap of demand and supply of professionals (Agarwal, 2012) [20]. The employees don’t consider BPO jobs to be long lasting. The very nature of BPO jobs is monotonous and mind numbing. These jobs are characterized by rotational shifts and long working hours. The pace of the work in call centers is controlled by automatic call distributor (ACD) or predictive dialer. The time spent on each call, break taken and response given is also monitored by team leaders. All these things create a pressure on employees leaving them more frustrated with the job and thus encouraging attrition (Budhwar, Varma, Singh and Dhar, 2007) [21]. Also, BPO doesn’t provide any scope for skill up gradation and the work experience of BPOs is not considered to be relevant for other job profiles (Babu, 2004) [22]. Due to such reasons, Indian outsourcing industry experiences high level of attrition rate and low level of employee tenure as employees keep looking for better opportunities continuously (Holman, Batt and Holtgrewe, 2007) [23]. BPO industry is said to have highest attrition rate and fast employment growth which reduces the average tenure of an employee in an organization. ITES/BPO witnesses the entry of highly qualified professionals but only for a short time period as

they make exit in short stint of time (Thite and Russell, 2010) [24]. Where engaged employees are the assets for an organization, disengaged employees act as liability. Disengagement of employees incurs huge cost to the companies. "Engagement" is not an individual phenomenon, it comprises many elements; job security, support from supervisors and availability of resources which act as positive predictors of engagement. Seijts and Crim (2006) [25] identified 10 C's of engagement: Connect, Career, Clarity, Convey, Congratulate, Contribute, Control, Collaborate, Credibility and Confidence. These include almost all aspects of engagement. According to Hickman (2010) [26] there are 5 dimensions of engagement: a) 'Belonging' wherein employees identify with the organization's core values and develop an understanding of relating their acts to the accomplishment of organization's goals b) 'Inspiration' which means that there must be a trust between the employee and the leader. Employees must know that their opinion counts, for this a participative form of leadership will work c) 'Understanding' entails communication of plans in time and clearly. Employees must understand what are they expected to do. Feedback practice should be followed to intimate employees about their performance d) 'Support' - an environment of supportiveness develops when supervisors are there for employees. It is the duty of leaders to develop such an organizational culture and e) 'Appreciation' means to encourage employees by appreciating their efforts in front of all, it gives a feeling of accomplishment and enhances morale.

It can be seen that according to many authors, an empowered work culture is capable of enhancing engagement level of employees which further boosts up the performance and productivity of employees. Such employees have fewer intentions to leave their organizations. BPOs implement strategic HR practices to empower their employees and enhance involvement. Still, the employees at these organizations are not engaged. ITES sector of India is renowned globally for rendering quality and relatively cheap services to the organizations. There is an urgent need to identify other factors which account for this attrition rate in BPOs

2.1 Women at BPOs

Women represent almost one-third part of the BPO sector and their contribution in this industry is major. Technology and certain regulatory changes have accentuated the demand of female employees in BPO sector. The Factories Act (2005) has been amended to allow women working in night shifts provided, all provisions in place to ensure their safety. Employers need to provide equal opportunities to women employees and must maintain their dignity at workplace (Pathak and Sarin, 2011) [1]. Major reasons behind joining this industry are easy entry and exit, fair remuneration, not much qualification required and lifestyle. Yet, they face serious issues like, night shifts, flexible working hours, transportation, socio-cultural factors and sexual harassment, which need to be addressed (Kaur and Gupta, 2012) [27]. Women working in night shifts suffer from certain behavioral and psychological

syndromes like, irritability, alcohol use, anxiety and depression etc. (Begum, 2013) [5]. In India, women are supposed to take care of their families as home maker which disturbs their work-life balance while working in this sector.

Gender discrimination is quite visible in this sector. Women employees need to put in their best effort to come at par with their male counterparts. Employers have less trust on them for accomplishing challenging jobs. Only repetitive tasks are given to them and that too without any autonomy. If any woman happens to precede as a boss then employers don't let her take important decisions independently (Pathak and Sarin, 2011) [1].

2.2 HR practices at BPOs

The prime objective of HRD is to unleash the unlimited potential of employees and to develop that potential by appropriate and systematic efforts (Selvi, 2012) [28]. HR department has begun to understand the issues faced by women now. Day care facilities for their kids, recreational activities, paid maternity leaves, provision for short leaves etc. have been initiated by the HR managers. Though these practices resolve this problem only at the face (Pathak and Sarin, 2011) [1]. Budhwar et. al.(2007) [21] opine that almost all BPO units follow their written HR strategies. The objectives of the department remain a) accurate 'recruitment' wherein they look for people having excellent communication, technical and problem solving skills, b) providing 'fair compensation which depends on grade, previous work experience, skills of employees, c) conducting 'performance appraisal' on the basis of how well an employee performed, d) giving 'training' required for performing a job like soft skills, basic software knowledge etc. and e) ensuring 'retention' by addressing key problem areas of BPOs like bureaucratic work settings, monotonous jobs, lack of growth opportunities etc. Practices followed by HR departments in Indian BPOs are setting benchmark for even developed nations, as they keep improvising their performance appraisal and career development programs. MNCs in India are adopting similar talent retention strategies as that of their western counterparts (Stumpf, Doh and Tymon, 2010) [29].

3. HYPOTHESES

H1: There exists a positive relationship between the level of engagement and level of empowerment of working professionals of BPOs.

H1a: There exists a positive relationship between belonging and the level of empowerment of working professionals of BPOs.

H1b: There exists a positive relationship between inspiration and the level of empowerment of working professionals of BPOs.

H1c: There exists a positive relationship between understanding and the level of empowerment of working professionals of BPOs.

H1d: There exists a positive relationship between support and the level of empowerment of working professionals of BPOs.

H1e: There exists a positive relationship between appreciation and the level of empowerment of working professionals of BPOs.

H2: There exists a significant difference between male and female employees of BPOs as regards their perceived level of empowerment.

H2a: There exists a significant difference between male and female employees of BPOs as regards their perceived level of autonomy in deciding job methods.

H2b: There exists a significant difference between male and female employees of BPOs as regards their perceived level of impact in organization.

H2c: There exists a significant difference between male and female employees of BPOs as regards their perceived level of performing meaningful job activities.

H2d: There exists a significant difference between male and female employees of BPOs as regards their perceived level of possessing complete knowledge of required skills

4. RESEARCH METHODOLOGY

This research initiative studies the perception of working executives of BPO sector in the NCR region regarding their levels of empowerment and engagement. The research has been carried out with a self constructed questionnaire. The questionnaire had several items related to the perception of respondents regarding belongingness, inspiration, understanding, support and appreciation extended to employees in corporate organizations. The questionnaire had two parts, part A was purely focused on collecting demographic details. Part B mapped the perception of respondents as regards different variables related to empowerment level of employees like, autonomy in deciding job methods and know how, importance of one's job, control over work etc. and engagement level of employees like autonomy in decision-making, growth opportunities, involvement, availability of information and communication etc. and The questionnaire was constructed on a five point Likert agreement scale to measure the responses on the decided variables. The questionnaire was subjected to review by experts and their inputs have been incorporated accordingly. Reliability of the same was computed to be Cronbach Alpha .83. According to Nunnally (1978, p. 245) the instruments used in basic research have reliability of about .70 or better.

The sampling was multistage. In the first place it was purposive wherein the researchers drew out a list of 32 BPO organizations in the NCR. The list was generated through the web link ([http://www.fundoodata.com/advance_search_results.php?and_new_industry_id\[\]=18andcompany_type_id\[\]=3andlevel_id=1andcity_id=0andcriteria=1andsearch_type=1](http://www.fundoodata.com/advance_search_results.php?and_new_industry_id[]=18andcompany_type_id[]=3andlevel_id=1andcity_id=0andcriteria=1andsearch_type=1)). *Eight companies were shortlisted from the list such that one company each was selected from the east ,west ,north , south ,central ,Gurgaon, Faridabad and Noida region .Questionnaires were administered to about 20 managerial level employees from each of these companies . About 100 completed questionnaires (out of 160) were taken for the study.* The data was subjected to inferential and descriptive analysis.

5. DATA ANALYSIS AND FINDINGS

The descriptive analysis of data is presented in Table 1. Engagement has been measured in five dimensions viz. Belonging, Inspiration, Understanding, Support and Appreciation.

Employees develop a sense of belongingness when they identify with the core values of their organization and start relating the impact of their acts on accomplishment of organizational goals. In BPOs, employees get easy entry but to sustain that job becomes a challenge. Rotational shifts, monotonous tasks and strenuous work environment make employees frustrated, disturb their work-life balance and thus, force them to move out of the organization. These jobs are not considered long lasting by the employees and job-hopping becomes a frequent practice. Owing to such reasons, employees are not able to establish that bond with the organization and don't develop a sense of belongingness towards their organization ($m=3.59$, $s.d.=0.58$). They feel that organization is concerned for its own profits and not towards their individual career goals. It de-motivates employees. Employees feel inspired when their leaders encourage them to participate in important decision making discussions and when they get fair remuneration according to their contribution. Relationship between team leaders and team members and remuneration plays a pivotal role here. These are the constituents of 'inspiration' dimension and the perception towards this dimension is lowest amongst BPO employees ($m=3.45$, $s.d.=0.66$). This is because top level management don't invite opinions of lower level employees. Also, BPO handles more than one process simultaneously where the concerned operation manager may discuss important and urgent things with the team leader only but not employees. Even if top or middle level managers don't show their concern, team leaders of the different processes take due care of the team members needs. They keep all communication channels in place in order to keep everyone aware, listen to problems faced by team members on workstations and maintain a healthy working environment. This forms the reason for enhanced level of perception of employees towards 'understanding' dimension ($m=3.88$, $s.d.=0.53$). The presence of such practices at workplace indeed enhances the intrinsic motivation of employees and creates a bond between employer and employees. Team leaders care for their team members and their professional needs. They act as liaison between them and top management. Furthermore, corporate programs and annual meets are organized to fulfill the social needs of employees as well as ensuring availability of required materials and information, so, many respondents are satisfied with the level of 'support' they receive from their organization ($m= 3.85$, $s.d.=0.51$). In order to enhance the empowerment level of employees, BPOs have developed a tradition to give "star of the week" and "employee of the month" titles to employees who outperform in the given tasks. These titles are announced amongst all employees and operations managers and gifts are given to such employees. Such recognition boosts up the morale of such employees and others as well. 'Appreciation' dimension of engagement comprises it all and employees are found to be

contented about it (m=3.81, s.d.=0.57). Overall Engagement of the professionals working in BPOs on the basis of these five dimensions is moderate (m=3.71 and s.d.=0.57). The reasons behind this are less concern of top management towards employees who are working day and night in different processes for the organization. Employees feel bounded as log-in hours and breaks are fixed. They are not free to move frequently out of their floors as punch-in entries are recorded for each employee.

Such reasons frustrate employees and force them to leave the organization.

	N	Minimum	Maximum	Mean	Std. Deviation
Understanding	100	2.25	5	3.88	0.53
support	100	2.75	5	3.85	0.51
appreciation	100	2	5	3.81	0.57
Belongingness	100	2	5	3.59	0.58
Inspiration	100	1.25	5	3.45	0.66
Engagement				3.71	0.57

Table 1: Mean scores of five dimensions of Engagement

In order to understand the relationship between the dimensions of engagement viz. belonging, inspiration, understanding, support, appreciation and the perceived level of empowerment, correlation coefficients were calculated. A higher coefficient indicates a stronger correlation between variables.

There existed a moderate and positive relationship between the perceptions of employees regarding the feeling of belongingness and perceived level of empowerment (r=.490, p<.01)- Table 2. Employees understand that in order to develop that intrinsic motivation to come to work, there has to be a “connect” between them and the organization. When employees feel that they are the integral part of their workplace, they feel empowered. The systems, practices and managers should give pertinence to this and use it to retain human capital. Therefore, H1a stands accepted.

		Belongingness	Empowerment
Belongingness	Pearson Correlation	1	.490**
	Sig. (2-tailed)		0
	N	100	100
Empowerment	Pearson Correlation	.490**	1
	Sig. (2-tailed)	0	
	N	100	100

Table 2: Correlation between Belongingness & Empowerment

A positive relationship was observed between the inspiration dimension and perception towards empowerment (r=.514, p<.01)- Table 3. This study was conducted on BPOs rendering IT services and it can't be denied that technology is dynamic in nature. In this sector, learning can take place only by exchanging information, so, team leaders and managers seek certain technical information and details of working by employees. They need to extend required support to such employees in return. Also, these team leaders seek others' opinions regarding certain decisions. It gives team members a sense of recognition in their department. On the remuneration front, overtime and incentive schemes also contribute to inspiration for working in this type of working environment. Hence, H1b is accepted.

		Inspiration	Empowerment
Inspiration	Pearson Correlation	1	.514**
	Sig. (2-tailed)		0
	N	100	100
Empowerment	Pearson Correlation	.514**	1
	Sig. (2-tailed)	0	
	N	100	100

Table 3: Correlation between Inspiration & Empowerment

Timely communication from team leaders, safe working environment and sound feedback mechanisms are the elements of understanding. It was observed that team members exhibit higher level of perceptions about the relationship of understanding and empowerment (r=.425, p<.01)-Table 4. This may be attributed to the reasons that BPOs have systems in place to provide regular and immediate feedback to the employees working in various processes. Every voice call or backend task gets vetted first by auditor and relevant feedback is given to the employee. Team leaders are informed and directed in advance to communicate all requisite information to the team members. Therefore, H1c also stands accepted.

		Understanding	Empowerment
Understanding	Pearson Correlation	1	.425**
	Sig. (2-tailed)		0
	N	100	100
Empowerment	Pearson Correlation	.425**	1
	Sig. (2-tailed)	0	
	N	100	100

Table 4: Correlation between Understanding & Empowerment

A positive relationship was observed between the support received from team leaders and peers and perceived level of empowerment ($r=.429, p<.01$)-Table 5. Role of leader support and peer relationship in enhancing empowerment can't be sidelined. Support from superiors is shown when they address the work related issues of employees and extend the required help. Relationship with peers plays an important role on and off workplace as it fulfills the social need of employees. BPOs frequently organize corporate meets and get-togethers to fulfill this need of employees and to make them aware of new trends of technology prevailing in global scenario in an informal manner. It was observed that presence of such provisions increases the perception of employees towards relationship of support and empowerment. So, H1d is accepted.

		Support	Empowerment
Support	Pearson Correlation	1	.429**
	Sig. (2-tailed)		0
	N	100	100
Empowerment	Pearson Correlation	.429**	1
	Sig. (2-tailed)	0	
	N	100	100

Table 5: Correlation between Support and Empowerment
 Appreciation, one of the most crucial dimensions of engagement is fairly addressed in BPO sector. Literature suggests that when employees receive recognition for their efforts and appreciated by team leaders and managers, they feel empowered. BPO employees possess positive perceptions about the relationship of appreciation and empowerment ($r=.441, p<.01$)-Table 6. BPOs have a culture to reward outperformers in front of all team members, team leaders and concerned managers. Titles like “Star of the week” and “Employee of the month” along with small gifts are given to such employees. This enhances the motivation level of employees and thus boosts up performance. Therefore, H1e is accepted.

		Appreciation	Empowerment
Appreciation	Pearson Correlation	1	.441**
	Sig. (2-tailed)		0
	N	100	100
Empowerment	Pearson Correlation	.441**	1
	Sig. (2-tailed)	0	
	N	100	100

Table 6: Correlation between Appreciation & Empowerment

Many authors suggest that empowerment is one of the major driving forces of engagement which further enhances

performance and productivity and helps to retain employees. BPO employees show a higher level of perception about the relationship between engagement and empowerment ($r=.705, p<.01$) from Table 7. The employers at BPO sector have understood the need of the hour. They look up to the standards of western nations in order to come at par with them. They have started decentralizing decision making powers to their employees in order to give autonomy to them, open door policy has been implemented by superiors in order to make employees feel comfortable contacting them, reward and recognition policies have been revived and a trend of corporate and annual meets have been started. As all five dimensions of engagement bear a positive and strong relationship with empowerment, H1 which states that there exists a positive relationship between engagement and empowerment is accepted. Literature suggested that empowerment and engagement are positively and strongly related in different sectors. No study has been conducted to establish this relationship in BPO sector empirically. The researchers of this study conclude that empowerment acts as a driving force for engagement in BPO sector as well and can work miraculously in retaining employees.

		Engagement	Empowerment
Engagement	Pearson Correlation	1	.705**
	Sig. (2-tailed)		0
	N	100	100
Empowerment	Pearson Correlation	.705**	1
	Sig. (2-tailed)	0	
	N	100	100

Table 7: Correlation between Engagement & Empowerment
 Levene's t-test for equality of variances (Table 8) has been applied to test hypothesis H2 that there exists a difference between male and female respondents as regards their level of empowerment in their organization. It can be observed that there exists no difference between male and female respondents as regards their perceived level of autonomy in deciding job methods ($t= -0.275, p<0.05$). This is because women constitute majority portion of the workforce of BPOs. HR managers prefer women employees as they possess required patient to handle queries, have convincing and soft voice and are good administrators. Recruiting female employees, for most of the processes remains priority for HR managers. Post joining, team leaders leave certain job related decisions up to them only, which gives women employees a sense of control over their work. Hence, H2a is rejected. The level of empowerment gets affected when one gets to know one's impact/influence in his/her organization. Employees feel empowered when they realize that their contribution to their organization matters.

There found to be no difference between male and female employees of BPO sector as regards their perceived level of impact in their organization ($t=0.739$, $p<0.05$). Women employees are very converse with the fact that they occupy a significant position in this industry because of traits they possess and so are the BPO managers. For their efforts in accomplishing given tasks they get recognition and appreciation by the managers. It enhances their motivation level and they feel that their contribution matters to the organization. Therefore, H2b stands refuted. Organizations in the present era are evolving as learning organizations and employees as knowledge workers. All employees are concerned towards what they do at their workplace. They want to enrich their job roles by adding some meaning to it. They demand challenging roles for personal growth. To meet such expectations of employees, HR department design tasks in specialized manner and thus, monotonous, repetitive tasks are being replaced by challenging ones. Women employees are given such challenging tasks to perform and they feel no difference is there in the kind of work they get to do and their male counterparts ($t= -0.63$, $p<0.05$). Females have overcome various taboos levied over them in past. They are mobile, professionally educated and are ready to take opportunities hands-on. Hence, H2c stands rejected. Information and skills required to perform such specialized tasks demand continuous upgradation. There found to be no difference between male and female employees of BPOs about possession of complete knowledge of required skills ($t= -0.183$, $p<0.05$). This can be because in order to garner the job related needs of employees, HR department of BPOs organizes workshops by experts. HR department in consolidation with other managers design policies to be at par with western MNCs. With such unbiased approach of managers in training employees, female employees believe that they possess complete knowledge of all required skills. Therefore, H2d is also refuted. It becomes very evident that HR department has understood the significance of female workforce very well and is working towards addressing their workplace issues. Owing to these findings, Hypothesis 2 which states that there exists a significant difference between male and female employees of BPOs as regards their perceived level of empowerment stands refuted.

6. CONCLUSION AND IMPLICATION

Indian ITES/BPO sector is rampantly growing but has its own set of challenges. Young professionals are entering the workforce in BPOs in huge numbers so; it has become crucial to understand their characteristics and expectations from workplace. Career development practices help managers in retaining talent at workplace (Sullivan and Mainiero, 2007) [30]. These practices can enhance the productivity and performance of employees. According to Dash, Singh and Vivekanand (2008) [31], chances of promotion and opportunities for personal growth also boost up the engagement level of BPO employees. Few good MNCs help employees in completing their post graduation in order to enhance their knowledge. It also motivates employees and makes them feel

empowered. They need people who have got required skills. Wrong selections also lead to enhanced employee turnover rate. Organizations must try to gain an insight into the behavioral dynamics of present workforce and its implications for the leaders of tomorrow. The analysis of the data tells that managers working in BPOs are not satisfied with the leadership shown by the immediate superiors as well as top management. 'Boundations' created by strict patterns of working, force employees to leave the companies. Only professional care is not enough for employees to keep them motivated. They are of the opinion that there is much to life than just work. Top management and superiors need to improve their personal working relationships in order to meet their individual needs. Top management must live up to their commitments in order to build trust amongst young managers. Also, the biasness must be eliminated completely from the system in order to make things transparent.

The BPO workforce which belongs to different backgrounds demands support, recognition and praise from their bosses. Female employees need to be empowered by employers. They are capable of performing challenging tasks, so must be provided with significant growth opportunities. They don't appreciate unethical or unjustified practices. Unfair allocation of workload, uncompetitive salary structure and unfair review system lessens their morale level leading to disengagement from organization. The jobs situations are needed to be redesigned. Advanced technologies like e-learning programs, simulation exercises, motivational training etc. must be used to train them. Challenging work, support system of bright and creative people, team membership, autonomy at workplace, continuous learning and concern for their personal needs can transform the face of working practices for females at BPOs. Family support is also required to enhance the mobility of women managers. Organizations take due care regarding the security of females working in BPOs. This must be accompanied with some autonomy at workplace as well. Stereotyping must be addressed seriously now by the means of mentoring and counseling male employees. Also, continuing education courses to keep such employees updated about the advancements of industry, career development skills training, opportunities for growth and promotion, mentoring and leadership development programs are other avenues which must be considered by the management (Monis and Sreedhara, 2011) [32]. These factors are important components of empowerment as well which make employees feel important in the organization.

7. LIMITATION OF THE STUDY AND AVENUES FOR FUTURE RESEARCH

There are certain limitations to this study. It was conducted on the sample of 100 respondents from the BPO organizations in Delhi-NCR. The results so obtained cannot be generalized for whole BPO sector. Time constraint to conduct this study posed another limitation.

BPO culture is characterized by monotonous tasks, rotational shifts, unfair distribution of rewards, stiff competition, stressful jobs and different working practices for male and female

managers. The employees of this dynamic industry are ambitious and look out for better opportunities always. This study can be extended to larger sample in order to identify other factors which affect performance of BPO employees. Career development and talent retention strategies can be studied in same sector by other researchers.

REFERENCES

- [1]. Pathak, S. and Sarin, A., "Management of Stress Among Women Employees in BPO Industry in India : A Contemporary Issue", *International Journal of Management and Business Studies*, Vol. 1, Issue 3, pp. 65-70, 2011.
- [2]. Desai, M., Majumdar, B. and Prabhu, G.P., " A study on employee engagement in two Indian businesses ", *Asian Journal of Management Research*, pp 81-97, 2012.
- [3]. Spreitzer, G.M., "Psychological empowerment in the workplace: Dimensions, measurement and validation", *Academy of Management Journal*, 38, 1442-1465, 1995.
- [4]. Mishra, P.S., "Increasing rate of attrition in BPOs", *Management and Labour Studies*, Vol. 32 No.1, pp. 7-21, 2007.
- [5]. Begum, K.J.A, "Women and BPOs in India", *International Journal of Humanities and Social Science Invention*, Volume 2, Issue 5, 2013.
- [6]. Raman, S. Raghu ,Budhwar, Pawanand Balasubramanian, G., "People management issues in Indian KPOs", *Employee Relations*, Vol. 29 Iss: 6 pp. 696 – 710, 2007.
- [7]. Kahn, W.A., "Psychological conditions of personal engagement and disengagement at work", *Academy of Management Journal*, Vol. 33, pp. 692-724, 1990.
- [8]. Towers Perrin, "Today: Understanding what drives employee engagement. The 2003 Towers Perrin talent report", Retrieved from http://www.towersperrin.com/tp/getwebcachedoc?webc=hrs/usa/2003/200309/talent_2003.pdf, accessed on July, 01, 2013.
- [9]. Devi, R.M., " Employee engagement is a two way street", *Human Resource Management International Digest*, Vol. 17, No. 2, pp. 3-4, 2009.
- [10]. Saks, Alan M., "Antecedents and consequences of employee engagement", *Journal of Managerial Psychology*, Vol. 21 Iss: 7, pp. 600 – 619, 2006.
- [11]. Pritchard, K., "Employee engagement in UK: meeting the challenges in public sector.", *Development and Learning in organizations*, Vol. 22, Iss: 6, pp. 15-17, 2008.
- [12]. May, D.R., Gilson, R.L. and Harter, L.M., " The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work", *Journal of Occupational and Organizational Psychology*, Vol. 77 No. 1, pp. 11-37, 2004
- [13]. Maslach, C., Schaufelli, W.B. and Leiter, M.P., " Job burnout", *Annual Review of Psychology*, Vol. 52, pp. 397-422, 2001.
- [14]. Kanooni, "The employee empowerment in Information Technology department", Retrieved from <http://drarmankanooni.com/research/employee+empowerment.pdf>, accessed on June 16, 2013.
- [15]. Lockwood, Nancy R, "Leveraging Employee Engagement for competitive advantage: HR's strategic Role", Retrieved from <http://www.shrm.org/research/articles/articles/documents/07marresearchquarterly.pdf>, accessed on June 23, 2013.
- [16]. Gallup Organization, "Employee Engagement: What's your engagement ratio?", Retrieved from consulting.gallup.com, accessed on February 11, 2013.
- [17]. Hallberg, U.E and Schaufeli, Wilmar B., "Same same but different? Can work engagement be discriminated from job involvement and organizational commitment?", *European Psychologist*, 11: 119-127, 2006.
- [18]. Stander, M.W., and Rothmann, S., " Psychological empowerment, job insecurity and employee Engagement", *SA Journal of Industrial Psychology*, 36(1), Art. #849, 2010.
- [19]. Elmuti, D., "Self-managed work teams approach: creative management tool or a fad?", *Management Decision*. MCB University Press. 35(3), 233–239, 1997.
- [20]. Agarwal, A.K, "The Scientific and Educational Bulletin. Health and Educational Millennium", No. 1(14T), 2012.
- [21]. Budhwar, P.S., Varma, A., Singh, V. and Dhar, R., " HRM systems of Indian call centres: an exploratory study", *The International Journal of Human Resource Management* , 17:5, pp. 881-897, 2007.
- [22]. Babu, Ramesh P. , "Cyber coolies in BPO – insecurities and vulnerabilities of non-standard work", *Economic and Political Weekly*, Vol. 39, No.05, pp. 492-497, 2004.
- [23]. Holman D, Batt R and Holtgrewe U., "The Global Call Centre Report: International perspectives on Management and Employment", Report of the Global Call Centre Network (US format). <http://www.ilr.cornell.edu/globalcallcenter/upload/GCC-Intl-Rept-US-Version.pdf>, accessed on August 13, 2013.
- [24]. Thite, M. and Russell B., "Work organization, human resource practices and employee retention in Indian call centers", *Asia Pacific Journal of Human Resources*, 48(3), pp. 356-374, 2010.
- [25]. Seijts, G.H and Crim, D., "What engages employees the most or, the ten C's of employee engagement", *Ivey Business Journal*, March/Apri, pp1-5, 2006.
- [26]. Hickman, Nikki, "The five dimensions of Employee Engagement", Retrieved from <http://texasceomagazine.com/departments/people-matters/the-five-dimensions-of-employee-engagement/>, accessed on February 18, 2013.

[27]. Kaur, Arvinder and Gupta, Shivani, "Women employees in BPOs and call centers: Problems and solutions(A case study of Chandigarh)", Gian Jyoti E-journal, Vol.1, Iss. 3, pp.80-92, 2012.

[28]. Selvi, K.R., "Human Resource Development Practices in BPO industries at Bangalore", Trans Asian Journal of Marketing and Management Research, Vol.1, Iss. 3-4, pp. 27-32, 2012.

[29]. Stumpf, S.A., Doh, J.P. and Tymon, W.G., " The strength of HR practices in India and their effects on employee career success, performance and potential", Human Resource Management, Vol. 49, No. 3, Pp. 353– 375, 2010.

[30]. Sullivan, S.E., and Mainiero, L.A., "Kaleidoscope careers: Benchmarking ideas for fostering family-friendly workplaces", Organizational Dynamics, Vol. 36, No. 1, pp. 45-62, 2007.

[31]. Dash, Mihir, Singh, Anubha and Vivekanand , " Motivation in ITES industry: dimensionality with reference to Herzberg's theory", ICFAI Journal of

[32]. Monis, S.H. and Sreedhara, T.N., "Employee satisfaction with career development Practices: a comparative study of indian and Foreign MNC BPO firms", Journal of Arts Science and Commerce, Vol.– II, Issue –1, January 2011.

Independent Samples Test								
		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Autonomy in deciding job methods	Equal variances assumed	0.062	0.804	-0.275	98	0.784	-0.04921	0.17879
	Equal variances not assumed			-0.278	47.432	0.782	-0.04921	0.17695
Impact in organization	Equal variances assumed	0.529	0.469	0.739	98	0.462	0.15119	0.20451
	Equal variances not assumed			0.71	43.153	0.482	0.15119	0.21298
Meaningful job activities	Equal variances assumed	0.075	0.784	-0.63	98	0.53	-0.12329	0.19559
	Equal variances not assumed			-0.59	41.292	0.558	-0.12329	0.209
Complete knowledge of skills required	Equal variances assumed	0.484	0.488	-0.183	98	0.855	-0.03399	0.18564
	Equal variances not assumed			-0.196	53.557	0.845	-0.03399	0.17331

Table 8: Difference in the level of empowerment of male and female working executives (t-test).

Group Statistics					
	gender	N	Mean	Std. Deviation	Std. Error Mean
Autonomy in deciding job methods	1	73	3.877	0.798	0.093
	2	27	3.926	0.781	0.150
Impact in organization	1	73	3.781	0.886	0.104
	2	27	3.630	0.967	0.186
Meaningful job activities	1	73	3.877	0.832	0.097
	2	27	4.000	0.961	0.185
Complete knowledge of skills required	1	73	3.781	0.854	0.100
	2	27	3.815	0.736	0.142

Table 9: Group Statistics- Empowerment (Male & Female employees)

Cloud Computing and Knowledge Management as a Service: A Collaborative Approach to Harness and Manage the Plethora of Knowledge

Meenu Dave¹, Mikku Dave² and Y. S. Shishodia³

Submitted in March, 2013; Accepted in August, 2013

Abstract - The advent of cloud computing has opened up numerous avenues which were not explored or harnessed in the best possible manner. Knowledge management is one of such fields. In order to identify, create, store, represent or disseminate knowledge, the existing tools, technologies or methodologies have not been able to achieve what an organization always desires for its progress. The under-utilized intellectual capital is not even accounted for properly due to improper implementation of knowledge management methodologies. The implementation and maintenance costs also play a major role in the unacceptability of the techniques at a vast scale. In order to implement cloud computing successfully, a wide range of knowledge and its simultaneous dexterous management is also a prerequisite. Enterprises, whether business or service, need to adopt such strategies, guidelines, and structures that implement this combo in an effectual manner. This paper discusses the different aspects like reasons for under-utilization of organizational knowledge, the intersecting issues of cloud computing and knowledge management, impact of cloud computing on knowledge management, benefits of Knowledge Management-as-a-service and major players in market providing Knowledge Management-as-a-service.

Index Terms – Cloud Computing, Knowledge Management, Knowledge Management-as-a-Service.

1. INTRODUCTION

Cloud computing has been in the market for quite some time. With the concept slowly settling in the world, the emergence or recognition of new cloud resources is making it much more viable as compared to that in the past. With the evolvement of the market, it is imminent to properly identify the available resources, and categorize the components of cloud computing according to the newly discovered patterns. Cloud computing when used, does not only have a higher impact on the technology; it greatly affects the people also.

Cloud computing describes the development of many existing technologies and approaches to computing into something different. Cloud separates application and information resources from the underlying infrastructure, and the mechanisms used to deliver them.

^{1,3} Dept. of Computer Science and Engineering, Jagan Nath University, Jaipur, India.

² School of Management, Poornima University, Jaipur, India.

E-mail: ¹meenu.dave@jagannathuniversity.org,

²mikkudave@gmail.com and ³pvc@jagannathuniversity.org

Cloud enhances collaboration, agility, scaling, and availability, and provides the potential for cost reduction through optimized and efficient computing. More specifically, cloud describes the use of a collection of services, applications, information, and infrastructure comprised of pools of computation, network, information, and storage resources. These components can be rapidly orchestrated, provisioned, implemented and decommissioned, and scaled up or down; providing for an on-demand utility-like model of allocation and consumption [1]. Service delivery models of cloud computing are increasing day by day. With the popularization of a few categories like Storage-as-a-service, software-as-a-service, Platform-as-a-Service, Database-as-a-service, Information-as-a-service, and Management-as-a-service, the attention has been shifted to one of the most important aspect of any organization, i.e. knowledge management.

Knowledge Management is a scientific process that initiates its working by amassing knowledge (both tacit and explicit), filtering it, structuring or restructuring it, storing and finally disseminating it. The dissemination process of the already stored knowledge is again very crucial, as it should also be in such a manner that

- The access to knowledge is timely, accurate and easy.
- The accessed knowledge aids adequately in decision making, and
- The available knowledge facilitates in creation or generation of new knowledge. [2]

A successful knowledge management strategy must address four key areas:

- knowledge management strategy as a core focus and competency,
- flexible structure for knowledge creation and dissemination,
- technology and processes, and
- skilled knowledgeable workers [3].

Knowledge management adds transparency to an otherwise opaque backdrop of the issues encompassing cloud computing because it provides the solution to the problems of the oversimplification, incompleteness and inadequacy of information. The need for knowledge management also arises because the processes, tools & technologies involved and adopted should be evolved and refined keeping in view the opportunities that cloud computing is offering.

2. REASONS FOR UNDER-UTILIZATION OF ORGANIZATIONAL KNOWLEDGE

Two reasons have been the core for the under-utilization of the organizational knowledge:

- The principal reason is the reluctance of the employees to divulge the tacit knowledge. An employee sees it either as a threat (because he fears losing his job on the grounds of competition), or finds the sharing of information a time-consuming and mundane work for which he does not gain any incentive.
- Secondly, most of the organizations lack the proper modus operandi, framework, tools and technologies to amass the freely flowing intellectual capital, which goes unused and totally wasted. The acknowledgement of the wasted capital and timely investment in accumulating this asset and there after proper utilization for decision making is not yet achieved on a large scale.

3. CLOUD COMPUTING AND KNOWLEDGE MANAGEMENT: THE INTERSECTING ISSUES

As technology changes, however, the need for a clear and definitive knowledge management is increasing. The increase in adoption of mobile devices, cloud computing, and virtual workplace only makes knowledge management more important [4].

The deployment and consumption modalities of cloud should be thought of not only within the context of 'internal' vs. 'external' as they relate to the physical location of assets, resources, and information; but also by whom they are being consumed by; and who is responsible for their governance, security, and compliance with policies and standards.

The risk for knowledge management-as-a-service also depends upon:

- The types of assets, resources, and information being managed
- Who manages them and how
- Which controls are selected and how they are integrated
- Compliance issues [5].

Key issues regarding data lifecycle security in the cloud with reference to knowledge management-as-a-service include the following:

- **Data security.** Confidentiality, Integrity, Availability, Authenticity, Authorization, Authentication, and Non-Repudiation.
- **Location of the data.** There must be assurance that the data, including all of its copies and backups, is stored only in geographic locations permitted by contract, SLA, and/or regulation.
- **Data remanance or persistence.** Techniques for completely and effectively locating data in the cloud, erasing/destroying data, and assuring the data has been completely removed or rendered unrecoverable must be available and used when required.
- **Data backup and recovery schemes for recovery and restoration.** Data must be available and data backup and recovery schemes for the cloud must be in place and effective in order to prevent data loss, unwanted data overwrite, and destruction.

- **Data discovery.** As the legal system continues to focus on electronic discovery, cloud service providers and data owners will need to focus on discovering data and assuring legal and regulatory authorities that all data requested has been retrieved.
- **Data aggregation and inference.** Practices must be in play to assure the data owner and data stakeholders that the data is protected from subtle "breach" when data is commingled and/or aggregated, thus revealing protected information (e.g., medical records containing names and medical information mixed with anonymous data but containing the same "crossover field") [1].

4. IMPACT OF CLOUD COMPUTING ON KNOWLEDGE MANAGEMENT

Cloud services are increasingly being used for the purpose of knowledge management because of the following reasons:

- Technological advancement related to ubiquitous high-speed internet connectivity
- Shrinking cost of data-storage,
- The propagation of smart mobile devices at electric speed around the world

These factors have helped in fulfilling the pre-requisite of simple, cost-effective and flexible information. The use of smart phones and tablets demonstrates the potential of cloud computing to empower the users with sophisticated and high-powered yet uncomplicated and easy-to-use computer applications and information, which was otherwise not so easy to access.

It is rare today to find a user who is not aware of Google and the data and information that lays only a few clicks away. Faced with this ever-growing repository of knowledge, organizations struggle to provide the right solution to customers faster than they can find it on the Internet. A few aspects of the Google-effect can leave organizations vulnerable. The source of information found on the Internet must be carefully evaluated by the user to ensure that it can support decision-making. Dissemination of corporate knowledge is slow and often does not meet the needs of the user population. It is important for corporations to develop a knowledge management strategy that responds to user needs in a timely and efficient manner – that takes information and data real time and provides that information and data in the right context from a trusted source – the corporation [4].

The cloud magnification effect has been felt as discussed below:

- **Maintaining freedom of information:** when internet communications were shut down during unrest in Egypt, a team of international companies combined cloud services and social networking to connect in-country mobile phone networks to deliver text messages internationally via Twitter, in just a couple of days [6].
- **Broadening frontiers of transformation:** the continuously falling costs of computing, data transmission and storage are

constantly broadening the opportunity for cloud services to transform businesses and business models

- Enabling social networking: it's important to note that the entire social networking phenomenon is enabled by cloud computing. It is believed that the eventual impact of social networking on businesses and the related improvement in productivity will be profound.
- Accelerating technology transitions: one big difference between the successful introduction of tablet computers in 2010 from earlier failed attempts is that cloud-based services are available today to do the background computation, storage and other "heavy lifting" that make tablet computers useful tools [5].

5. BENEFITS OF CLOUD COMPUTING FOR KNOWLEDGE MANAGEMENT PERSPECTIVE

Keeping in view, the knowledge management process, the following requirements can be catered efficiently with cloud computing:

- Cloud computing greatly reduces the technology related costs. Synergies are created through reduction in computing resources, time and modelling process. The large and variable storage needs for maintaining huge knowledge management systems can be accomplished by the use of cloud storage models.
- It provides access to variety of services to the users.
- It opens gates to several options for the knowledge users, earlier unknown to them.
- Cloud computing is an effective way to streamline the knowledge and make it readily available also.
- It accelerates the development and acquiescence of the competencies and capabilities of knowledge works in any organization.
- It extends the use of open-source services and shared developments across the globe.
- It reduces the activities and costs related to management of infrastructure.
- The work structure is changing these days, where requirement of access to information could be anywhere and at anytime. Cloud resources are the best answer to such indispensability.
- Through cloud computing, one can gain faster access to technology and the process of knowledge management becomes more connected and collaborative, as the structure is framed on highly scalable and reliable web oriented architecture.
- It improves access to data that corroborates the decision-making process and consolidates the studies & research processes. Latest project status, innovative ideas, new experiences and finest practices can thus be shared within an enterprise. It even motivates the employees for bringing forth creative strategies while carrying out the regular work, and enjoying it as well.
- It enhances awareness about the processes as the experiences are gained in different regions, sectors & fields of the world.

- Search patterns in knowledge management systems may vary and may experience sudden fluctuations and hiccups and can be better managed with an elastic/scalable computing model.
- Majority of knowledge management activities are generic in nature for which SaaS applications may be provided.
- Enterprises refrain from employing knowledge management tools and techniques primarily due to the cost incurred, but with cloud computing, pay-as-you-go concept can help solve this problem, as the required and limited services or resources are available at hand.
- Access to knowledge can be allowed on the basis of user's permission level.
- With the aid of cloud computing, knowledge can be converted into an asset which acts as a stimulant for innovations and research.

6. MAJOR PLAYERS IN MARKET PROVIDING KNOWLEDGE MANAGEMENT AS A SERVICE

Cloud based services, in context to the knowledge management platform are increasing day by day. Many players have entered the arena, and are providing assorted services. A few are discussed as follows:

Salesforce Knowledge Management as a Service

Knowledge management software as a service from salesforce.com makes it easy to capture, review, catalog, and access information, so agents can find what they need when they need it. Knowledge management system by Salesforce includes these key features:

- A solution encyclopedia
- A private and public knowledge base
- Intuitive browsing and searching
- Solution administration
- Solution workflow
- Multilingual solutions
- Suggested solutions
- Solution ratings [7].

BMC Knowledge Management as a Service

BMC provides a powerful knowledge content search engine that not only helps service desk analysts find solutions to incidents, but also provides users with access to resources for resolving their own issues. BMC Knowledge Management as a Service is part of the BMC Remedy IT Service Management Suite. This offering delivers cloud-enabled service desk, self-service, inventory, and basic knowledge management capabilities hosted on the Cloud Platform [8].

Office 365 and SharePoint Online

Office 365 allows with anywhere access to email, web conferencing, documents, and calendars. It includes business-class security. Office 365 provides easy-to-use, cloud-based management tools in a single location. Through the administration user interface IT staff can set up new user accounts, control access to features, and see the status of all Office 365 services and tools in real time.

Microsoft SharePoint Online, which is part of the Office 365 offering, provides out-of-the-box facilities for implementing

Knowledge Management as a Service for most enterprise needs. Microsoft SharePoint Server technology delivered as an online service makes the power of the cloud work for the organization. SharePoint Online helps in creating sites to share documents and information with colleagues and customers [9].

Igloo Software

Igloo is a web-based platform for collaborating at work. It's a digital workplace that enables to share files, find answers, solve problems, locate information and expertise and tap into the collective knowledge of customers, partners and peers, virtually anywhere.

- Communicate with integrated messaging, presence & status
- Share, organize & manage documents & multimedia
- Collaborate through blogs, forums, wikis, polls and events
- Build rich user profiles to locate expertise & knowledge
- Find what you need faster through activity streams & search [10].

SpringCM Solutions

SpringCM offers powerful content cloud services platform available for businesses today. The service provided makes it easy to share documents, collaborate around content, streamline business processes, and deliver better business outcomes. A broad range of information - from content management, document and file sharing to collaboration, automated workflows, and cloud technology - in a wide range of formats are provided [11].

7. CONCLUSION

The adoption of cloud computing through knowledge management route will enable organizations to gain velocity with respect to information technology and to cater according to the continuously changing requirements of the market. It will empower not only the enterprises but also all the key participants.

Cloud providers must ensure that the information security systems they provide are responsive to customer requirements and the data, both primary and secondary, must be preserved as authentic and reliable. Expansion and up gradation is imperative for cloud computing because if such an improvement is done in procedures or techniques or policies even for one customer then it is going to be advantageous for others as well.

REFERENCES

- [1]. "Security Guidance for Critical Areas of Focus in Cloud Computing V2.1", Cloud Security Alliance, December 2009. Available online at: <http://www.cloudsecurityalliance.org/guidance/csaguide.v2.1.pdf>
- [2]. M. Dave, M. Dave and Y.S. Shishodia, "Knowledge Management and Organizational Competencies: A Harmonic Collaboration", International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), Vol. 2, No. 12, pp. 45-50, December 2012.
- [3]. R. C. Neto, R. R. Souza, J. T. Neves, R. R. Barbosa, "Strategic knowledge management: In search of a knowledge-based organizational model", *Comportamento Organizacional E Gestão*, Vol. 14, No. 2, pp.247– 256, 2008.
- [4]. J. L. Mohr, "Knowledge Management as a Strategy: Driving a Competitive Edge Through Knowledge", Mind the IT Gap, LLC, 2012. Available online at: http://www.servicemanagementfusion.com/conference/~media/HDICConf/2012/Files/White%20Papers/Julie_Mohr_KM_Strategy.pdf
- [5]. "Cloud Computing: Issues and Impacts", Global Technology Industry Discussion Series, Ernst & Young, 14 April 2011. Available online at: [http://www.ey.com/Publication/vwLUAssets/Cloud_computing_issues,_impacts_and_insights/\\$File/Cloud%20computing%20issues%20and%20impacts_14Apr11.pdf](http://www.ey.com/Publication/vwLUAssets/Cloud_computing_issues,_impacts_and_insights/$File/Cloud%20computing%20issues%20and%20impacts_14Apr11.pdf)
- [6]. "Cloud, smart computing to drive tech sector growth", Forrester, Business Line (The Hindu), via Dow Jones Factiva, © 2010 The Hindu Business Line, 16 December 2010. Available online at: <http://www.thehindubusinessline.in/2010/12/16/stories/2010121653090700.htm>
- [7]. "Knowledge Management Tools that Work the Way You Do", Available online at: <http://www.salesforce.com/crm/customer-service-support/knowledge-base-system/>
- [8]. "Knowledge Management Solutions: Create and share IT know-how", Available online at: <http://www.bmc.com/products/offering/Knowledge-Management.html>
- [9]. "Microsoft Office 365 and SharePoint Online", Available online at: <http://www.microsoft.com/en-in/office365/sharepoint-online.aspx>
- [10]. "Igloo Software: The new digital workplace", Available at: <http://www.getapp.com/igloo-software-application>
- [11]. "Springcm Resources", Available at: <http://www.springcm.com/services>

Additional Fault Detection Test Case Prioritization

Ritika Jain¹ and Neha Agarwal²

Submitted in April, 2013; Accepted in August, 2013

Abstract - Regression testing is used to confirm that previous bugs have been fixed and that new bugs have not been introduced. Thus regression testing is done during maintenance phase and applied whenever a new version of a program is obtained by modifying an existing version. To perform a regression testing a set of new test cases and old test cases that were previously developed by software engineers are reused. This test suite is exhaustive in nature and it may take long time to rerun all test cases. Thus regression testing is too expensive and the number of test cases increases stridently as the software evolves. In present work, an additional fault detection test case prioritization technique is presented that prioritizes test cases in regression test suite based on number of concealed faults detected by test cases. Both noncost cognizant and cost cognizant prioritization of test cases have been performed using proposed technique and efficiency of prioritized suite is assessed using APFD and APFDc metric respectively.

Index Terms – APFD, APFDc, Cost-cognizant, Regression testing, Test case prioritization.

1.0 INTRODUCTION

A software product goes through requirement elicitation phase, design phase, implementation phase and testing phase before being fully operational and ready for use [1]. At the coding time developers often saves the set of test cases to test the module written and to reuse them later whenever software undergoes changes. When testing phase starts testing team creates a separate test suite, test environment and test oracle to validate the whole software system against customer satisfaction. Most of the time it is assumed that the software life cycle end after its delivery. But a long lasting maintenance phase starts after the deployment of the software to the customer's site. During this long life software evolves through numerous modifications and additions based on faults, change of user requirements, change of working platform or environments, change of government policies, and so forth. With the evolution of software product, maintaining its quality becomes more difficult and harder due to its numerous released versions, which goes on incrementing with every new set of changes to the previous version. These modifications into the existing software or addition of new features to the software may create new faults or may cause it to work improperly.

Thus assuring the quality of software product along with these modifications and additions is the challenging task of maintenance phase. Sometimes the quality may become worse than before. On the other hand, users hope that the new version of the software should be easy to use, has more features, and has better quality than before.

Regression testing is used to confirm that fixed bugs have been fixed and that new bugs have not been introduced. Regression testing refers to that portion of test cycle in which a program P' is tested to ensure that not only does the newly added or modified code behaves correctly, but also the code that carried over unchanged from the previous version P continues to behave correctly. Thus regression testing is done during maintenance phase and applied whenever a new version of a program is obtained by modifying an existing version. Regression testing is sometimes referred to as "program revalidation". The term "corrective regression testing" refers to regression testing of a program done after making corrections to the previous versions. Another term "progressive regression testing" refers to a regression testing of a program done when new features are added to the previous version. To perform a regression testing a set of new test cases and old test cases that were previously developed by software engineers are used. This test suite is exhaustive in nature and it may take long time to rerun all test cases. Thus regression testing is too expensive and the number of test cases increases stridently as the software evolves.

Researchers [2-3] have provided effective regression testing techniques. The simplest one is to reuse all test cases that were run before the modification of previous version of the software. But it might waste time and resources due to execution of unnecessary tests. Therefore it is desirable to run only a subset of test suite, which can be chosen by using regression test selection techniques. The drawback of test subset selection techniques is that some important and crucial test cases might be evaluated as worthless and might not be selected for execution, which might cause some effected portion of the software to be remained untested. Another approach is to permanently remove the number of test cases from the test suite by eliminating redundant test cases and thus reducing the test suite size, which can be accomplished by using test suite reduction techniques. The downside of this is that it might degrade the fault detection capability with the reduction of test suite size. The above discussed problems can be solved by test case prioritization techniques, in which test cases in a test suite are rearranged in an attempt to make sure that faults get revealed at earlier stage of testing process. Test case prioritization techniques schedule test cases in an order such that those with higher priority, according to some objective criteria, are executed before than those with lower priority, to meet some performance goal. Test case prioritization can be

^{1, 2}Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University, Noida-201303, India
E-mail: ¹jritika27@gmail.com and ²nagarwal2@amity.edu

used in conjunction with test subset selection and test reduction techniques to prioritize test cases in selected or minimized test suite [4-6].

In this paper, a new test case prioritization algorithm has been proposed that prioritizes the test cases in the test suite based on the maximum number of faults detected which has not been revealed by any other test case executed so far. In first experiment, test cost and fault severities are considered as unity and then test cases are prioritized using proposed algorithm. The efficiency of prioritized suite is then measured using APFD metric. In second experiment, test cost and fault severities are incorporated while prioritizing test cases and measured using APFDc metric. In the end experimental results are analyzed and compared.

The rest of this paper is structured in the described format. In Section 2, a brief description of test case prioritization technique is given followed by literature survey in Section 3. The new algorithm to test case prioritization is presented in Section 4. The experimental studies along with the obtained results and their analysis are given in Section 5. Finally, some conclusions are drawn in Section 6.

2. TEST CASE PRIORITIZATION

Test case prioritization problem is defined as finding a permutation of test cases to maximize an objective function, which reflects the prioritization goal. The formal definition of this problem can be expressed as follows [3, 7-9].

2.1 Problem Statement

Given: T, a test suite; PT, the set of permutations of T; f, a function from PT to the real numbers.

Problem: Find $T' \in PT$ such that $(\forall T'') (T'' \in PT) (T' \neq T'') [f(T') \geq f(T'')]$

Here, PT represents the set of all possible prioritizations (orderings) of T, and f is a function that, applied to any such ordering, yields an award value for that ordering. The definition assumes that higher award values are preferred over the lower ones.

2.2 Measuring Efficiency

In order to measure the effectiveness of test case ordering obtained through test case prioritization technique, a metric APFD (Average Percentage of Fault Detected) has been defined in the literature [3]. This metric measures average percentage of faults detected (rate of fault detected) against average percentage of test suite executed. Its value ranges from 0% to 100%. Higher APFD value reveals higher rate of fault detection and lower value reveals lower rate of fault detection. This metric provides a mean to compare efficacies of various prioritization techniques. In a formulaic presentation of the APFD metric, let T be a test suite containing n sequential test cases, and F be a set of m faults revealed by T. Let T' be some ordering of T. Let TF_i be the first test case in T' that reveals fault i. The APFD for T' is:

$$APFD = 1 - \frac{TF_1 + TF_2 + TF_3 + \dots + TF_n}{nm} + \frac{1}{2n} \tag{1}$$

The limitation of APFD metric is that it treats all test cases with equal cost and all faults with equal severity. However, practically these factors possess distinct value for different test case and drastically affect the effectiveness of test case ordering. In such cases, the APFD metric provides unsatisfactory results. To triumph over the weakness of APFD metric a new metric APFDc (Average Percentage of fault Detected per Cost) has been suggested in the literature [3]. This metric measures average percentage of test case cost incurred against average percentage of fault severity detected. The APFDc metric can be quantitatively described as follows: let T be a test suite containing n test cases with cost t_1, t_2, \dots, t_n . Let F be a set of m faults revealed by T, and let f_1, f_2, \dots, f_m be the severities of those faults. Let TF_i be the first test case in an ordering T' of T that reveals fault i. The weighted (cost-cognizant) average percentage of faults detected during the execution of test suite T' is given by the equation:

$$APFD_c = \frac{\sum_{i=1}^m (f_i \times (\sum_{j=TF_i}^n t_j^{-\frac{1}{2}} t_{TF_i}))}{\sum_{j=1}^n t_j \times \sum_{i=1}^m f_i} \tag{2}$$

3. LITERATURE SURVEY

Researchers [7, 9-10] have provided variety of test case prioritization techniques. Most of these techniques prioritize test cases based on their coverage information such as statement coverage, branch coverage, loop coverage, function coverage, condition coverage, and fault detected. Different prioritized suite of test cases are then obtained based on different coverage techniques and evaluated using APFD metric in order to contrast them. However, these code coverage based prioritization techniques have two limitations.

First, these techniques consider the value of test cost and fault severity as equivalent. On the contrary, factors allied to test cost and fault cost radically affect the ordering of test cases in prioritized suite and so its efficacy. As a consequence, researchers work out different ways to compute test cost and fault cost and provide cost cognizant test case prioritization techniques [2-3, 8, 11] that include these factors while prioritizing test cases. The efficacy of cost cognizant test case prioritization techniques can be measured through APFDc metric.

Literature [12-13] also proposed historical value based approach for cost cognizant test case prioritization in which a historical value model is used to calculate historical value of test cases based on the historical test case cost and the historical fault severities.

Second limitation is that, testers needed source code to assess coverage information of test cases and if source code is not available then applying this technique will be very difficult. In order to beat this difficulty, history-based test prioritization techniques were proposed and studied [14-15] that uses historical execution data of test cases in order to determine their priority. In addition, some researchers [16-17] had

employed genetic algorithm to history-based test case prioritization. They collect data such as test cost, fault severities, and detected faults of each test case from the latest regression testing and apply genetic algorithms to acquire better prioritized suite.

4. PROPOSED TECHNIQUE

In present communication, a new regression test suite prioritization algorithm is presented, that prioritizes the test cases in the regression test suite with the goal of finding the maximum number of faults at the early phase of the testing process. It is presumed that the desired execution time to run the test cases, the faults they reveal, and the fault severities are known in advance. This technique considers total number of concealed faults detected by each test case, to prioritize them. First the total number of faults detected by each test case is found. From this, the one which detects maximum faults is selected and then the fault coverage data of each unselected test case is adjusted to indicate their new fault coverage data (faults detected by each test case that are not yet discovered). Then the test case which covers maximum faults is selected. If there is more than one test case which covers maximum number of faults then choose them randomly and again adjust the fault coverage data of unselected test cases. This process is repeated until all faults have been covered. When all faults have been covered, same process is repeated for the remaining test cases. To make this technique accustomed to the situation where test costs and fault severities vary, instead of summing the number of new faults covered by a test case t to calculate the worth of t , the number of new faults f covered by t is multiplied by the criticality-to-cost adjustment $g(\text{criticality}_t; \text{cost}_t)$ for t . The notion behind the use of this computation is to reward those test cases that have greater cost adjustments when weighted by the additional faults they cover.

Algorithm 1: when test cost and fault severity are unity

Input: Test suite T , fault trace history F_h
 Output: Prioritized test suite T'

- 1: begin
- 2: set T' empty
- 3: initialize values of vector F_{cov} as "uncovered"
- 4: While T is not empty do
- 5: for each test case $t \in T$ do
- 6: calculate total number of faults f covered by t using F_h and F_{cov}
- 7: end for
- 8: select test t that cover maximum number of faults
- 9: append t to T'
- 10: update F_{cov} based on the faults covered by t
- 11: if all faults has been covered
- 12: initialize values of vector F_{cov} as "uncovered"
- 13: end if
- 14: remove t from T
- 15: end while
- 16: end

Algorithm 2: when test cost and fault severity are different

Input: Test suite T , fault trace history F_h , test criticalities T_{crit} , test costs T_{cost}

Output: Prioritized test suite T'

- 1: begin
- 2: set T' empty
- 3: initialize values of vector F_{cov} as "uncovered"
- 4: While T is not empty do
- 5: for each test case $t \in T$ do
- 6: calculate total number of faults f covered by t using F_h and F_{cov}
- 7: calculate award value of t as $f * g(\text{criticality}_t; \text{cost}_t)$ using T_{crit} and T_{cost}
- 8: end for
- 9: select test t with the greatest award value
- 10: append t to T'
- 11: update F_{cov} based on the faults covered by t
- 12: if all faults has been covered
- 13: initialize values of vector F_{cov} as "uncovered"
- 14: end if
- 15: remove t from T
- 16: end while
- 17: end

F_{cov} is a vector having values "covered" or "uncovered" for each fault in the system. The vector record the faults that have been covered by previously selected test cases. T_{crit} is the criticality of test case measured by summing the severities of all faults covered by test case.

4.1 Comparative Techniques

No prioritization. As an experimental control, one comparator technique that we consider is no prioritization, where no prioritization technique is applied and test cases are executed in sequential order.

Random ordering. Another comparator technique that we consider is the random ordering of the test cases in the test suite.

4.2 Estimating Test Cost and Fault Severity

For cost cognizant prioritization it is required to obtain cost of each test case and severity of faults each test case reveal. [3, 13] makes available some measures to compute test cost and fault severity. Test costs are greatly diversified in software testing. Depending on the criteria, a test cost can be computed over several factors such as machine time, human time, test case execution time, monetary value of the test execution, and so forth. Similarly, fault severity can also be measured by depending upon criteria such as test criticality (the criticality of the test case that detects a fault) and function criticality (the criticality of the function in the code that is covered by the test case). In our approach, test cost is refined as the test case execution time of a test case and Fault severity is refined to test case criticality, which is devoted to each test case by summing up the severities of each fault that test case reveal.

4.3 Award Value Calculation

In cost cognizant prioritization the test cases are prioritized based on award value. The award value of a test case is calculated using the formula

$$\text{Award value} = f \times g(\text{criticality}_t, \text{cost}_t)$$

Where criticality_t is the total severity of faults detected by test case t , cost_t is the cost of test case t , and g is a function that maps the criticality and cost of t into a value. (Function g simply divides criticality_t by cost_t). Greater the award value more will be the chances of the test case to be selected for execution.

5. RESULTS AND DISCUSSION

Table 1 shows test cost and fault severity of test cases and faults respectively [11]. The proposed technique is examined on data given in Table 1 and comparative analyses are drawn.

new fault coverage data of remaining test cases adjusted to T8 are shown in Table 2.

Fault / test case	T1	T2	T3	T4	T5	T6	T7	T9	T10
F2		\$	\$		\$				
F3				\$		\$			\$
F4		\$	\$						
F7				\$	\$		\$		
F8	\$					\$			
F9				\$		\$			\$
Number of faults	1	2	2	3	2	3	1	0	2

Table 2: Fault coverage data of test cases adjusted to test case T8

Now there are two test cases that expose three new faults. Next test case to be executed is T4. In this way, the prioritized suite is produced by applying the additional fault detection technique. Table 3 shows test suite prioritized order both for comparative techniques and proposed technique.

Prioritization technique	Prioritization order
No prioritization	T1-T2-T3-T4-T5-T6-T7-T8-T9-T10
Random ordering	T5-T3-T9-T1-T8-T6-T2-T10-T7-T4
Additional fault detection	T8-T4-T2-T1-T6-T3-T9-T5-T10-T7

Table 3: Prioritization order based on additional fault detection without considering cost

The efficiency of this prioritized order is measured through APFD metric and its value is given in Table 4.

Prioritization technique	APFD (%)
No prioritization	53
Random prioritization	60
Additional fault detection	75

Table 4: APFD Value of prioritized suite

It is seen that proposed prioritization technique increases the rate of fault detection capability of regression test suite upto 75% from 53% when there is no prioritization. The APFD value of prioritization order obtained through proposed technique is also greater than random ordering which makes it clear that random prioritization is never reliable. The proposed technique is also compared with the technique given by Kavitha and Sureshkumar [11]. It can be presented that the APFD value obtained by proposed technique is 75% whereas it is reported to be 70% for the same test data by Reference [11]. APFD graphs of unprioritized suite, random ordered suite, and additional fault detection prioritized suite is demonstrated in Figure 1a, 1b, and 1c respectively. The horizontal line in the graph represents average percentage of test suite executed and

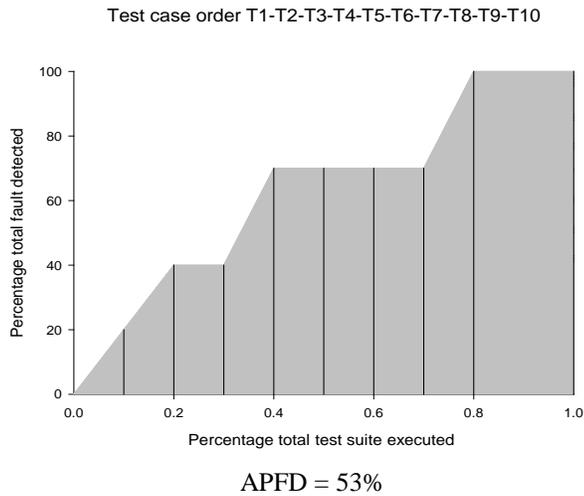
Fault / test case	T 1	T 2	T 3	T 4	T 5	T 6	T 7	T 8	T 9	T1 0	Fault severity
F1								\$	\$		6
F2		\$	\$		\$						6
F3				\$		\$				\$	6
F4		\$	\$								10
F5								\$			8
F6								\$	\$		10
F7				\$	\$		\$				4
F8	\$					\$					20
F9				\$		\$				\$	12
F10	\$							\$			6
Number of faults	2	2	2	3	2	3	1	4	2	2	
Time	9	8	14	9	12	14	1	10	0	0	13

Table 1: Time taken to find out the fault and the severity value [11]

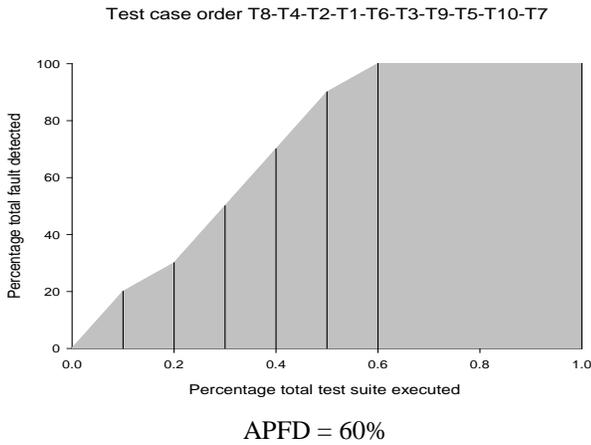
5.1 Experiment 1

Test cases are prioritized based on the total number of concealed faults detected by each test case. Here test cost and fault severity is considered as unity. From Table 1 as it can be seen, test case T8 has revealed maximum number of faults, thus the first test case to be executed in prioritized suite is T8. The

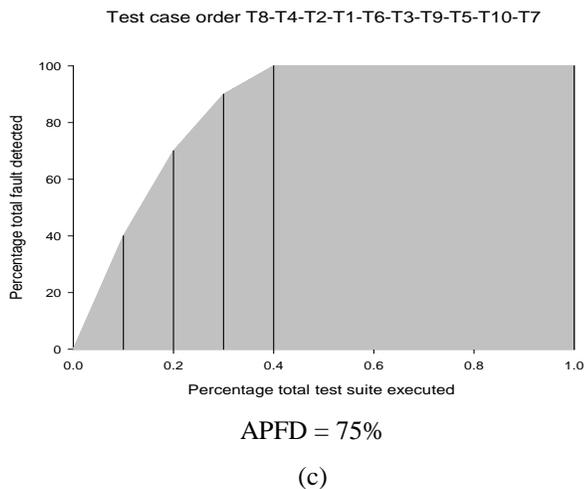
vertical line represents average percentage of total faults detected.



(a)



(b)



(c)

Figure 1: APFD graph of (a) unprioritized test suite, (b) random ordering test suite and (c) test suite from proposed technique

5.2 Experiment 2

Since test cost and fault severity greatly varies for each test case and fault, ignoring them can never produce appropriate and satisfactory results. Thus test cost and fault severity is integrated with test cases and faults respectively to produce a prioritized suite that detects more and more severe faults by incurring less cost to execute test cases. Test cases are prioritized based on award value of each test case. Efficiency of the prioritized test suite is measured through APFDc metric. Prioritized suite that is obtained by applying algorithm 2 on the comparative techniques and on data shown in Table 1 is presented in Table 5.

Prioritization technique	Prioritization order
No prioritization	T1-T2-T3-T4-T5-T6-T7-T8-T9-T10
Random ordering	T5-T3-T9-T1-T8-T6-T2-T10-T7-T4
Additional fault detection	T8-T6-T2-T4-T1-T9-T10-T3-T7-T5

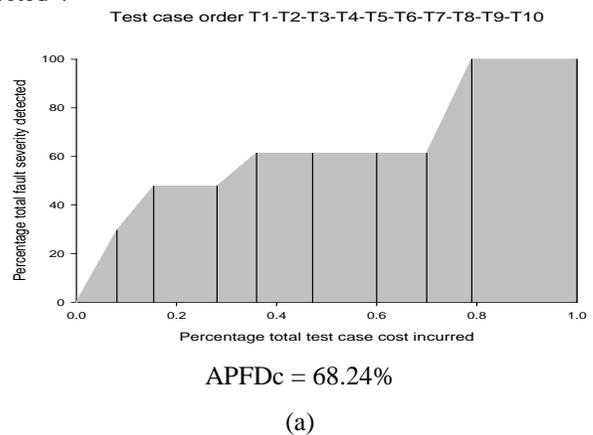
Table 5: Prioritization order based on additional fault detection considering cost

For these prioritized order their efficiency is measured through APFDc metric and its value is given in Table 6.

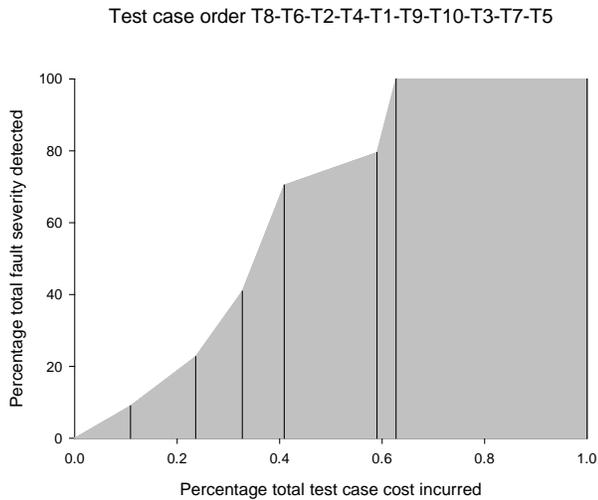
Prioritization technique	APFDc (%)
No prioritization	68.24%
Random prioritization	63.95%
Additional fault detection	85.64%

Table 6: APFDc value of prioritized suite

It can be observed from Table 6 that the efficiency of random ordering can sometimes be less than unprioritized test suite. Proposed prioritization technique shows highest rate of fault exposing potential (85.64%). It is also observed that cost cognizant test case prioritization has considerable higher fault exposing rate than noncost cognizant prioritization technique. APFD graph of unprioritized suite, random ordered suite, and additional fault detection prioritized suite is demonstrated in Figure 2a, 2b, and 2c respectively. The x-axis represents weighted “average percentage of test cost incurred” and y-axis represents weighted “average percentage of fault severity detected”.

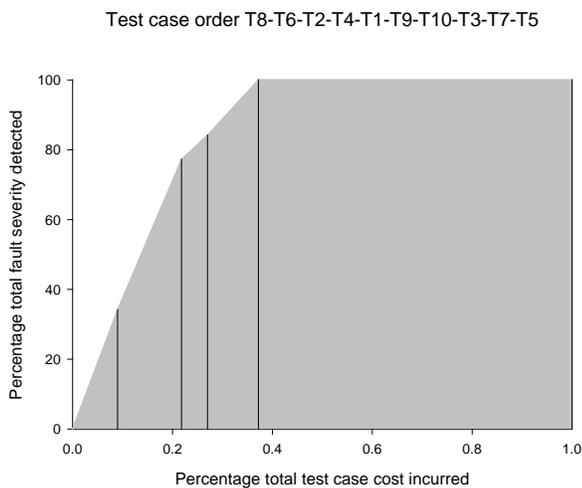


(a)



APFDc = 63.95

(b)



APFDc = 85.64

(c)

Figure 2: APFDc graph of (a) unprioritized test suite, (b) random ordering test suite and (c) test suite from proposed technique

6. CONCLUSIONS

In present communication, a new prioritization technique for Regression testing is presented that prioritizes test cases based on exposure of undetected faults by each test case to improve the rate of fault detection. The results show that the proposed technique leads to improve the rate of detection of severe faults at early stage in comparison to nonprioritized order and random order of test cases. When test cases are prioritized without considering the cost of tests and severity of faults, prioritized suite of proposed technique gives 75% APFD value, which is very large as compared to no prioritization and random

prioritization. When the same technique is integrated with test cost and fault severities, prioritized order of proposed technique gives 85.64% APFD value which is not only larger than comparative techniques but also larger than APFD value of noncost cognizant prioritized suite of proposed technique. It is also observed that the number of test cases required to find all faults is less in case of proposed prioritization technique. The important feature of this technique is that it exposes abundance amount of severe faults in short duration of test suite execution.

REFERENCES

- [1]. K.K. Agarwal, and Y. Singh, "Software engineering," IIIrd ed, New age international publishers, 2008.
- [2]. P.R. Srivastava, "Test case prioritization," *Journal of Theoretical and Applied Information Technology*, vol. 4 (3), pp. 178-181, 2008.
- [3]. A.G. Malishevsky, G. Rothermel, and S. Elbaum, "Cost-cognizant test case prioritization," *Technical Report TR-UNL-CSE-2006-0004*, Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, Nebraska, U.S.A., 12 March 2006.
- [4]. K.K. Aggrawal, Y. Singh, A. Kaur, "Code coverage based technique for prioritizing test cases for regression testing," *ACM SIGSOFT Software Engineering*, vol. 29, no. 5, Sep. 2004.
- [5]. R. Malhotra, A. Kaur, Y. Singh, "A regression test selection and prioritization technique," *Journal of Information Processing Systems*, vol.6, no.2, June 2010.
- [6]. G. Rothermel and M.J. Harrold, "A safe, efficient regression test selection technique," *ACM Trans. Software Engineering and Methodology*, vol. 6, no. 2, pp. 173-210, Apr. 1997.
- [7]. G. Rothermel, H. Roland H. Untch, Mary Jean Harrold, "Prioritizing Test Cases For Regression Testing", *IEEE Transactions on Software Engineering*, vol. 27, no. 10, October 2001.
- [8]. S. Elbaum, A. Malishevsky, G. Rothermel, "Incorporating Varying Test Costs and Fault Severities into Test Case Prioritization," *Proc. 23rd International Conference on Software Engineering*, May 2001.
- [9]. S. Elbaum, A. Malishevsky, G. Rothermel, "Test Case Prioritization: A Family of Empirical Studies," *IEEE Transactions on Software Engineering*, vol. 28, No. 2, February 2002.
- [10]. G. Rothermel, R. Untch, C.chu, M. Harrold, "Test case prioritization:an empirical study," *Proc. International conference on Software Maintenance*, pp. 179-188, Aug. 1999.
- [11]. R. Kavitha, N. Sureshkumar, "Test Case Prioritization for Regression Testing based on Severity of Fault", *International Journal on Computer Science and Engineering*, vol. 02, No. 05, 2010, 1462-1466.
- [12]. H. Park, H. Ryu, J. Baik, "Historical value based approach for cost cognizant test case prioritization to improve the effectiveness of regression testing," *2nd*

- International Conference on Secure System Integration and Reliability Improvement*, pp. 39-46, July 2008.
- [13]. P. Srivastava, "Performance Evaluation of Cost-cognizant Test Case Prioritization", *UACEE International Journal of Computer Science and its Applications*, Volume 1, Issue 1.
- [14]. A. Khalilian, M. A. Azgomi, Y. Fazlalizadeh, "An improved method for test case prioritization by incorporating historical test case data", *Science of Computer Programming*, volume 78, issue 1, 1 November 2012, pages 93-116.
- [15]. J. M. Kim, A. Porter, "A history based test prioritization technique for regression testing in resource constrained environment," *In Proc. of the International Conference on Software Engineering*, Orlando, USA, pp. 119-129, 2002.
- [16]. Y.C. Huang, C.Y. Huang, J.R. Chang, T.Y. Chen, "Design and Analysis of Cost-Cognizant Test Case Prioritization Using Genetic Algorithm with Test History," *IEEE 34th Annual Computer Software and Applications Conference*, 2010.
- [17]. Y.C. Huang, K.L. Peng, C.Y. Huang, "A history-based cost-cognizant test case prioritization technique in regression testing," *The Journal of Systems and Software*, 85, 2012, 626– 637.

Short-Term Spectral Feature Extraction and Their Fusion in Text Independent Speaker Recognition: A Review

Ruchi Chaudhary

Submitted in March, 2013; Accepted in August, 2013

Abstract - The paper gives an overview of Text-independent short-term-feature-extraction methods of Speaker Recognition System, for clean as well as noisy environment and their fusion at different levels. The basics of extracting feature, which is an imperative component for speaker recognition system, have been discussed along with their variants. The evolution of the conventional methods to 'State-of-the-Art' feature extraction methods are also brought out. This review helps in understanding the developments, which have taken place at various stages, with their relative merits and demerits. A comparative study of different techniques has been done at the end of each section to justify the choice of techniques available in the 'State-of-the-Art' speaker recognition systems. This study quantifies the effectiveness of short-term features for speaker identification.

Index Terms - Short-term feature Extraction, Speaker Recognition, Mel-frequency-Cepstral-coefficients (MFCC), Fusion.

1. INTRODUCTION

Speaker Recognition is the process of automatically recognizing the speaker by the use of the vocal characteristics [1-4]. 'State-of-the-Art' speaker recognition system uses number of voice characteristics, which include physical difference of the vocal production organs (shape of vocal tract, larynx size), and the manner of speaking (accent, rhythm, annotation style, pronunciation, pattern choice, vocabulary etc.) [1-4]. Fundamentally, speaker recognition process involves, extraction of speaker's specific characteristics (called features) from the given speech samples (*process known as feature extraction*) and the speaker model is trained and stored into the system database. In the recognition mode, the feature vector is extracted from the unknown's person's utterance and compared against the trained model. The purpose of feature extraction stage is to extract the speaker-specific information called feature vectors, represent the speaker-specific information based on one or more of the following: vocal tract, excitation source and behavioral tracts. All speaker recognition systems use set of scores to enhance the probability and reliability of the recognizer.

National Technical Research Organisation,
Old JNU Campus, Block-III, New Delhi-110067.
E-mail: ruchimakani@rediffmail.com

Speaker recognition systems can be divided into text-dependent and text-independent recognition systems [1-4]. In text-dependent speaker recognition systems, the speaker uses the same phrase at the time of enrolment/verification and in text-independent speaker recognition systems, which are more challenging and complex, the text of the speech at the time of enrolment/verification are completely random in nature. Success, in both recognition tasks, depends on extracting and modeling the speaker specific characteristics of the speech signal.

Pre-processing plays a vital role in speaker recognition as it decreases the acoustic effect and channel-mismatch. It is considered good practice to reduce the amount of variation in the data that does not carry important speech information. In other words, the pre-processing removes all non-relevant information such as background noise, characteristics of recording device etc. Voice activity detection (VAD), pre-emphasis filtering, normalization and mean subtraction are the few widespread commonly used steps in pre-processing [1-4]. By applying a pre-emphasis filter the glottal waveform and lip radiation characteristics are eliminated. The following fig. 1 shows the basic block diagram of speaker recognition system.

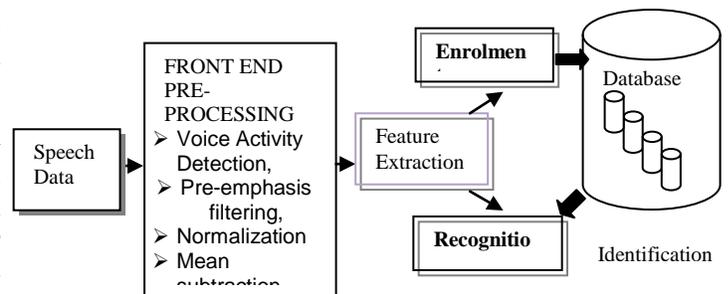


Figure 1: Block diagram of Speaker Recognition System

In order to have a good recognition performance, the front-end of the recognizer should provide feature vectors that can capture the important characteristics of an utterance. Besides, the front-end should also demonstrate reasonable performance in robust environment. Features can be categorized into (1) Short-term spectral features (2) Voice source features (3) Spectro-temporal features (4) Prosodic and high level features. Short-term spectral features are also referred as low-level features, have been dominantly used for speaker identification, as they are easy to compute and yield reasonably good performance, because these reflect information about the speaker's physiology and do not rely on the phonetic content

(which makes them inherently text-independent) [2]. Short-term analysis has been effective because of the quasi-stationary property of speech. The higher-level features also have the potential of increased robustness to channel variation, since phonetic, prosodic, and lexical information usage or temporal patterns do not change with the change of acoustic conditions [2]. Long-term information refers to features that are extracted over a longer region than a short-term spectral feature frame. Prosodic features capture variations in intonation, timing, and loudness that are specific to the speaker [2]. Because such features are supra-segmental i.e., extend beyond one segment, they can be considered a subset of long-term features.

The fundamentals of short-term features extraction methods used for speaker recognition systems, in clean and noisy environment, are discussed in section II and the overview of the fusion methods used in short-term-spectral-feature-based speaker identification system is provided in section III. Finally, section IV presents the conclusions of the study.

2. SHORT-TERM FEATURES EXTRACTOR

Generally, in any short-term feature extraction technique, framing and windowing are required before the mathematical computation of feature vectors. The general block diagram of short-term feature extraction is shown in the fig. 2.

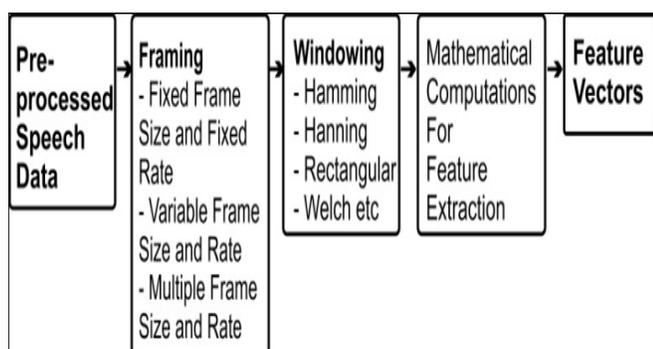


Figure 2: Block diagram of Short term Feature extraction

2.1 Framing

In short-term spectral feature (frame-based processing) are computed from frames size of about 20-30 ms in duration. Within this interval, the signal is assumed to remain stationary and a spectral feature vector is extracted from each frame. Generally, the frame shift is fixed to about half of the frame length. The *Fixed Frame size and Rate (FFFR)* approach results into some problems such as accidentally meeting noise frames. It may not be the best choice because characteristics of speech signal may rapidly change, especially at phonetic boundaries. The recognition accuracy increases, if the frame interval is directly controlled using phonetic information [5-11]. To avoid accidentally meeting noise frames problems, *Variable Frame Size and Rate (VFSR)* technique was proposed by *Qi eng Zhu et. al.* [5], which select optional frame size and frame rate depending on speaker rate, to capture sudden changes in spectral information with time. *Ponting et. al.* [6]

and *Samkwong et. al.* [8] showed that VFSR could successfully improve the performance of speech recognition; however, it increases the burden in identifying the spectral changes in speech. The study in [9] demonstrated that the spectral variations in speech can also be captured by combining multiple frame size (MFS) and multiple frame rate (MFR). It was shown that combined MFS and MFR gives better performance, compared to fixed frame size for language and speech identification task. *Multiple frame size and rate (MFSR)*, proposed by *H S Jayanna et. al.* [9] shows that MFSR framing method increases the performance of speech recognition, as it can generate more number of feature vectors, in case of limited training/testing speech data. *Chi-Sang Jung et. al.*[10], in 2010, based on experiments, have shown that feature frame selection methods result, in minimum redundancy within selected frames, and maximizes the relevancy to speaker recognition characteristics which produces consistent improvement.

2.2 Windowing

The window functions such as rectangular, hamming, hanning, welch etc, which minimize the spectral distortion, are needed because of the finite-length effect of the Discrete Fourier Transform (DFT). An overlap between the windows is needed since there is loss of information at the borders of window. In speech processing a hamming window is mostly used as it gives more precise frequency estimation [12]. The global shape of the DFT magnitude spectrum known as spectral envelope contains information about the resonance properties of the vocal tract which is highly relevant for speaker identification. A simple model of spectral envelope uses a set of band-pass filters to do energy integration over neighbouring frequency band. In 1969, Fast Fourier Transform (FFT) which is a fast implementation of DFT, based cepstral coefficients, was used in study [13].

2.3 Feature Mathematical Computation Methods

To develop robust speaker identification system, it is necessary to understand the different feature mathematical computation methods for extracting the features from the speaker's speech. A good feature set should represent all the components of speaker information, however, there does not yet exist globally 'best' feature but the choice is a trade-off between speaker discrimination, robustness, and practicality.

S Pruzansky, et. al. [14-15], conducted the first speaker identification study in 1963. In his study he had shown that spectral energy patterns and their variance yielded good performance for the speaker recognition. *Glenn* [16] in his study, in 1967, suggested that acoustic parameters produced during nasal phonation are highly effective for speaker recognition and average power spectral of nasal phonation can be used as the features for the speaker recognition. In 1969, Fast Fourier Transform (FFT) based cepstral coefficients were used in speaker verification study [4].

In 1969, *Atal* [17] demonstrated the use of variations in pitch as a feature for speaker recognition. Other acoustic parameters

such as glottal source spectrum slope, word duration and voice onset were proposed as features for speaker recognition by Wolf [18] in 1972. The concept of Linear Prediction (LP) for speaker recognition was introduced by Atal in 1974 [19]. The basic idea behind linear predictive analysis is that a specific speech sample at the current time can be approximated as a linear combination of past speech samples [4]. Through minimizing the sum of squared differences (over a finite interval) between the actual speech samples and linear predicted values, a unique set of parameters or predictor coefficients can be determined [4].

In 1976, *Sambur et. al.* [20-21] proposed the use of orthogonal linear prediction coefficients (LPC) as feature in speaker identification. LPC analyze the speech signal by estimating the formants, removing their effect from speech signal, and estimating the intensity and frequency of the remaining buzz. The process of removing the formants is called inverse filtering, and the remaining signal is called residue [21]. In 1977, long term parameter averaging, which includes pitch, gain and reflection coefficients for speaker recognition was studied [22] and, it was shown that reflection coefficients are informative and effective for speaker recognition system. *Reflection Coefficients*, defined as a sequence of ratios of the discontinuity of the cross-sectional area of the vocal tract, various improved derivatives of LPC i.e. Line Spectral frequency (LSF) [23], log Area Ratio (LAR) [24], Perceptual log Area Ratio (PLAR) [24], Perceptual Linear Prediction (PLP) [25] etc., were also studied by the researchers. Perceptual Linear Prediction (PLP) works by warping the frequency and spectral magnitudes of the speech signal based on auditory perception tests.

Mermelstein [26] conducted experiments in 1976, and showed that the cepstral coefficients are useful for representing consonantal information. Cepstral coefficients are the result of a cosine transform of the real algorithm of the short-term energy spectrum [27]. A study carried by *Rosenberg et. al.* [20] suggested that adjacent cepstral coefficients are highly correlated and hence all coefficients may not be necessary for speaker recognition. The LPCs are converted to cepstral coefficients using autocorrelation techniques on linear frequency scale, called *Linear Predictive Cepstral Coefficients* (LPCCs). *Mel frequency cepstral coefficient* (MFCC) were introduced in 1980, which use mel-frequency scale, are popular features in speech processing [4, 31-34]. The advantage of mel-frequency scale is that it approximates the nonlinear frequency resolution of the human ear. As with any filter bank based speech analysis technique, an array of band pass filter is utilized to analyze the speech in different frequency bandwidths. In MFCC parameterization, the position of the band pass filter along with the linear frequency scale is mapped to Mel-scale according to equation (1):

$$f_{\text{mel}} = 2595 \log_{10}(1 + f/100) \quad (1)$$

Overlapping of triangular filters in low frequency region of the energy spectrum (upto 1 KHz) in MFCC, are closely spaced.

While smaller number of less closely spaced triangular filters, are used to cover the high frequency zone, to weight the DCT of the speech so that the output is approximately of the same order as the energies of the filter bank signals. The experiments conducted by researchers, have shown that as the number of filters in the filter-bank increases, the identification rate of system increases [33-39]. In [41], Gaussian filters (GF) are also suggested, as they improved the system performance over the conventional triangular filter. GF provide much smoother transition from one sub-band to the other preserving most of the correlation between them. The means and variances of these GFs can be independently chosen in order to have control over the amount of overlap with neighbouring sub bands.

A study by *Reynolds et. al.* [28], in 1994, compared the different features like MFCCs, LPCCs, LPCs and Perceptual Linear Prediction Cepstral Coefficients (PLPCCs) for speaker recognition. They reported that among these features, MFCCs and LPCCs gave better performance than other features. At present, even though various alternative features are available, the MFCC seem to be difficult to beat in practice.

In 1981 *Furui* [35] introduced the concept of dynamic features, to track the temporal variability in feature vector, in order to improve the speaker recognition performance. In addition to short-term frame energy, formant transitions and energy modulations also contain useful speaker-specific information. A common way to incorporate some temporal information to features is through 1st and 2nd order time derivative estimates, known as delta (Δ) and double delta (Δ^2) coefficients, respectively. A study made by *G R Doddington* in 1985 [36], converts the speech directly in to pitch, intensity and formant frequency and these features were also demonstrated to provide good performance.

Md. Sahidullah, et. al. [37], have proposed inverted filter bank structure, such that the higher frequency range is averaged by more accurately spaced filters and a smaller number of widely spaced filters are used in the lower frequency range. This feature set named as *Inverted Mel Frequency Cepstral Coefficients (IMFCC)* follow the same procedure as normal MFCC, but using reversed filter bank structure. The combination of both enhances accuracy [38-39] and as such they can be considered complementary.

Another approach using Wavelet Transform instead of Discrete Cosine Transform, in the feature extraction stage, was proposed by Nengheng Zheng, *et.al* [40] in 2007. According to this, WOCOR is derived by wavelet transformation of the LP residual signal and is capable of capturing the spectro-temporal properties of vocal source excitation. There has been an array of these features used such as wavelet filter banks. It has been shown that the speaker identification system outperform when combination of both MFCC and WOCOR are used as feature extraction.

2.4 Features Extraction Methods for Noisy Environment

Robust speech techniques attempt to maintain the performance of a speech processing system under diverse conditions of operation (environmental differences between training/testing

conditions). To improve the performance of speaker recognition systems in noisy environment, approaches can be roughly divided into three categories, namely, *robust speech feature extraction*, *speech enhancement* and *model-based compensation* for noise, as shown in fig 4.

In the case of *speech enhancement*, some initial information about speech and noise is needed to allow the estimation of noise and clean up of the noisy speech. Widely used methods in this category include spectral subtraction (SS) and Wiener filtering. Statistical models such as Hidden Markov Models (HMMs), *Parallel Model Combination* (PMC), *Vector Taylor Series* (VTS) and *Weighted Projection Measure* (WPM) are generally classified into model-based compensation category [41-46]. The compensation techniques try to remove the mismatch between the trained models and the noisy speech, to improve the performance of the system.

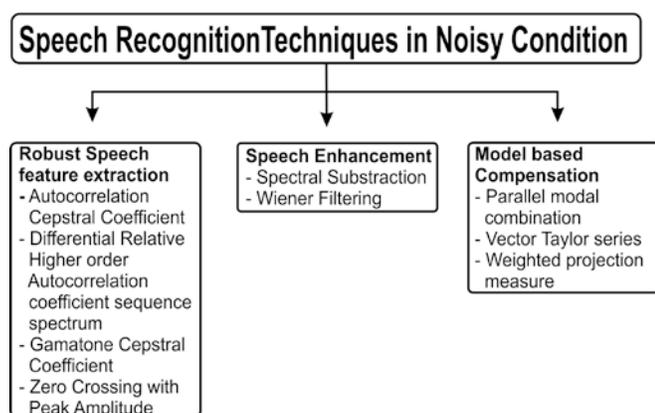


Figure 4: Speech Recognition techniques in noisy environment

LPCC, PLP and MFCC, perform well in clean environment, however, they suffer severe performance degradation in noisy conditions, especially when there is a noise level mismatch between the training and testing environments [41-46]. One of the major issue with these methods is that they are very sensitive to additive noise. Use of, *robust speech feature extraction* methods for improvement in speaker identification system in noisy environment is discussed in following paragraphs.

Use of the autocorrelation domain in speech feature extraction has recently proved to be successful for robust speech recognition [46]. A temporal filtering procedure on the autocorrelation sequence has been proposed [47] to minimize the effect of additive noise, which is derived based on filtering the temporal trajectories of short time one sided autocorrelation sequence. This filtering process minimizes the effect of additive noise which are stationary in nature at low SNR's, white, F16 and factory noise and no prior knowledge of noise characteristics is required. Results of experiments conducted at [47-48], indicate that *an Autocorrelation Mel Frequency cepstral coefficient (A-MFCCs)* significantly improves the performance of speaker identification system in noisy environment.

Poonam Bansal, *et. al.* [49] in 2010 evaluated the speech recognition performance of the *Differentiated Relative Higher Order Autocorrelation Coefficient Sequence Spectrum* (DRHOASS) features. DRHOASS uses only the higher-order autocorrelation coefficients for spectral estimation and discards the lower order autocorrelation coefficients. Speech coefficients, using the magnitude spectrum of the relative one-sided higher-order autocorrelation sequence, differentiating it and then processing it through a Mel filter bank, were finally parameterized in terms of MFCCs. It is shown in [49] that DRHOASS features perform almost similar to MFCC features, for clean speech; however, it performs better than the MFCC features for noisy speech. It was found that higher order autocorrelation coefficients along with additional filtering improved the robustness of the speech recognition system under different background noises.

Gammatone Cepstral Coefficients (GTCC) and *Zero-Crossings with Peak Amplitude* (ZCPA) are also claimed to have better performance than the conventional algorithms, especially in extremely noisy conditions (<15dB SNR) [50, 51]. GTCC, an acoustic feature extraction based on an auditory filterbank realized by Gammatone filters, was introduced for large vocabulary speech recognition [50]. Gammatone filters were used to characterize data obtained by reverse correlation from measurements of auditory nerve responses. The zero-crossing analysis (ZCA) of speech waveforms, proposed by *Doh-Suk Kim, et. al.* [51] has advantages over autocorrelation, power spectrum and linear prediction methods. This is because of the reason that these conventional methods data extraction by sampling a time waveform, depends on the maximum frequency content in the time signal whereas, ZCA requires a number of extracted samples, determined by the average rate of zero-crossing intervals. In ZCPA, the reciprocal of time intervals between two successive zero crossings are collected in frequency histograms from which frequency information is extracted.

In summary, spectral features like band energies, formants, spectrum and cepstral coefficients represent mainly the speaker specific information due to the vocal tract. Excitation source feature like pitch, variation in pitch information from LP residual and glottal source parameters represents mainly the speaker specific information due to vocal cord. Long-term, features like duration, intonation, energy, AM and FM components represents mainly the speaker specific information due to the behavioural traits. Practically use of MFCC & LPCC as short term feature extraction method for speaker specific information provides accuracy and reliability. However, GTCC, ZCPA, DRHOASS provide better performance in noisy environment.

3. FUSION OF COMPLEMENTARY INFORMATION

Recent researches have indicated that the appropriate fusion is an approach to meet stringent performance requirement of a recognition system. The fusion hold the promise of improving basic recognition accuracy by adding complementary information, not captured by the conventional features alone

and, possibly, robustness to acoustic degradations from channel and noise effects, to which majority of the features are highly susceptible. Fusion reduces the cost of implementation by using several cheap sensors rather than one expensive sensor [52-57, 60-61].

A common recognition system includes fingerprint, face, hand geometry, finger geometry, iris, retina, signature, voice, gait, smell, keystroke, ECG, etc. Recognition systems based on single source of information are called unimodal systems. Since, the unimodal system alone may not be able to achieve the desired performance requirement in real world applications, the use of multimodal biometric authentication system, which combines information from multiple modalities increases the performance [52]. The scope of this section is to cover the fusion preferred for unimodal short-term-feature-based speaker identification system in order to enhance the performance system.

The performance of the information fusion system is highly dependent on the effectiveness of the fusion technique implemented. By considering pre-defined data attributes, including channel characteristics and speaker's emotional and stress patterns detectable in conversations, the fusion method need to be fine-tuned to improve results. Over a period of time, numbers of information fusion techniques have been proposed for speaker identification systems. Generally, the information fusion can be done at: (i) feature level, (ii) score level, or (iii) decision level [55]. The block diagram of fused system (different level of fusion) is shown in fig 5. In subsequent paragraphs, fusion strategy and various fusion levels methods for unimodal short-term-feature-based speaker identification system are discussed and the comparison in term of their performance, is also brought out.

combination schemes such a product, sum, minimum, maximum, median, average etc., have been utilized and their performance compared empirically. It is found that the sum rule outperforms the other combination schemes and it is most resilient to estimation errors [57-58].

3.2 Feature Level Fusion

In feature fusion, multiple features are concatenated into a large feature vector, and thereafter, a single model is trained with these fused large feature vector [59]. Each feature vector need to be normalized before concatenated. A well-known data fusion strategy in speaker identification system is to concatenate the cepstral vectors with their delta and delta-delta cepstral into a long feature vector. Also the fundamental frequency has been used in addition with the cepstral vectors to improve recognition accuracy. Researchers have used various accent features such as pitch, energy, intonation, MFCCs, formants, formant trajectories, etc., and some have fused several features to increase accuracy as well. In [59], a fusion of MFCC, accent-sensitive cepstral coefficients (ASCC), delta ASCCs, energy, delta energy and delta-delta energy was done to improve the accuracy of the speaker identification system. In commercial biometric systems, the MFCC, Delta Coefficients and formant feature fusion method are preferred over the other fusion techniques.

3.3 Score Level Fusion

In score level fusion, each different feature set is processed separately with specialized classifier, and thereafter, output scores obtained from each classifier are combined. Each of the different features sets acts as an independent "expert", giving its opinion about the unknown speaker's identity. The *fusion rule* then combines the individual experts match scores. This approach is referred to as *score-level fusion*. Score level fusion is generally preferred in systems since the matching scores relatively easy to obtain and contain sufficient information to make genuine and impostor case distinguishable. The score mean and variance of multiple non-target speaker models are used to normalize the score of the target speaker models. Since the scores generated by identification system can be either similarity scores or distance scores, one needs to convert these scores into a same nature. The common practice, which is followed, is to convert all the scores into similarity scores.

Researchers at [37-39], used MFCC and IMFCC feature vectors, with two separate Gaussian Mixture Model (GMM) classifiers and fused their scores. Likewise, same principle has been adopted for Gaussian filter based MFCC and IMFCC also [37-39]. In both cases, for each speaker, two scores were generated; improved system performances were observed after fusion in both the cases. In [59], two support vector machine (SVM) classifiers, using both MFCCs and LPCCs separately, were fused to achieve high accuracy of the system.

In 1995 *P. Thevenaz et. al.* [62] reported that the energy of LP residual alone, gives less performance and combining it with LPCC improves the performance as compared to that of LPCC alone. Several other studies have been reported that combining

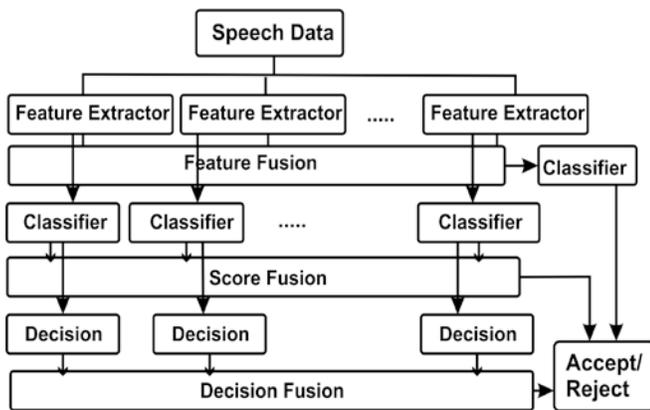


Figure 5: Block diagram of Fused System

3.1 Fusion strategy

In course of time, various architectures and schemes have been directed by researchers for combining multiple features/classifiers. Authors at [57-58] developed a common theoretical framework for combining features or classifiers which use distinct pattern representations. Number of possible

LP residual with MFCC improves the performance as compared to that of MFCC alone [63]. In 1996, Plumpe developed a technique for estimating and modeling the glottal flow derivative waveform from speech for speaker recognition [64]. In this study, the glottal flow estimate was modelled as coarse and fine glottal features, which were captured using different techniques. Also it was shown that combined coarse and fine structure parameters gave better performance than the individual parameter alone. In 2003, B Peskin, et. al. [65] reported that combination of prosodic features like long-term pitch with spectral features provided significant improvement as compared to only pitch features. Nengheng Zheng, et. al. [40] in 2007 studied that the complementary contributions of MFCC and WOCOR in speaker identification, significantly outperforming the one using MFCC only. Thereafter, a score level fusion technique was used for combining MFCC and WOCOR for speaker identification, in order to improve the performance. Further, the system comprised of a fusion of classification scores from adapted HMM and GMM, where scores from two recognition systems were fused [66].

The area of automatic speaker recognition has been dominated by systems using only short-term, such as cepstral features. These systems generally produced low error rates. Recently published works have demonstrated that such high-level information can also be used successfully in automatic speaker recognition systems by improving accuracy and potentially increasing robustness. Wide ranging high-level-feature-based approaches using pronunciation models, prosodic dynamics, pitch gestures, phone streams, and conversational interactions were explored. In [66], vocal tract and vocal cord parameters are integrated to identifying speakers. In spite of the two approaches having significant performance differences, the way they represent speech signal is complementary to one another. In [67-68] have also shown great improvement in speaker verification accuracy through fusion of low and high speech levels classifiers. However, the complexity of the system increases.

3.4 Decision Level fusion

Each of the expert classifier produces an identified speaker label, and the fusion rule combines the individual decisions e.g. by majority voting, called *decision level fusion* strategy. In other words, it is hardening the decisions of the individual classifiers. The inclusion of multi-resolution classifiers enhances fusion capabilities in handling noisy patterns, thus increasing accuracy. Other approaches to combine classifiers (decision fusion) include the rank-based methods such as the Borda count, the Bayes approach, the Dempster-Shafer theory, the fuzzy integral, fuzzy connectives, fuzzy templates, probabilistic schemes, and combination by neural networks [69-76].

The results at [61] show that the four acoustic feature based subsystems outperform the tokenization subsystems and the best subsystem is the LPCC-SVM system. Comparisons with the MFCC-GMM system, the temporal discrete cosine transform (TDCT)-GMM system captures the longer time

dynamic of the spectral features but it requires more training data.

Author at [76] has shown that the decision fusion based on the output of the GMM and SVM classifiers increases the discriminative power, as does fusion between classifiers based on spectral features and classifiers based on prosodic information.

3.5 Comparison on Fusion Methods

Although feature-level fusion may improve recognition accuracy, it has several short comings. Firstly, fusion becomes difficult if a feature is missing (e.g. F0 of unvoiced sounds) or the frame rates of the features are different. Secondly, the number of training vectors needed for robust density estimation increases exponentially with the dimensionality. This phenomenon is known as the *curse of dimensionality*. It can be seen that the feature-level fusion improves the performance over the individual classifier in the case of MFCC and its delta features. However, in all other cases it degrades the performance [77]. Decision level fusion is considered to be rigid due to the availability of limited information. However, this is the best fusion strategy, when all feature sets are used. Features using parametric linear prediction based estimation of the spectral envelope (LPCC variants) provide the best speaker recognition results and are the most useful in fusion, followed by minimum variance distortion-less response based estimation. The score level fusion gives the best results in all cases fusing feature. It can be seen that the feature-level fusion improves the performance over the individual classifier in the case of MFCC and its delta features. Thus, fusion at the score level is usually preferred, as it is relatively easy to access and combine the scores presented by the different modalities. In [78], score and decision fusion for the mel-cepstrum and corresponding delta features and demonstrated that the score level fusion performed consistently better. Furthermore, they observed that the computational complexity for the feature fusion is higher than that of the score and decision level fusion. Hybrid feature based speaker identification system has been proposed in [79], and various combination of the feature-level, score-level and decision-level has been observed to give the advantage on over the single level fusion methods.

For any text independent short-term-feature-based speaker identification system, fusion at the score level is generally considered appropriate due to the ease in accessing and consolidating matching scores.

4. CONCLUSIONS

In this paper, short-term feature extractions techniques are discussed for text-independent speaker recognition system in clean as well as in noisy environment and also including various fusions preferred in order to maximize system performance for speaker identification system.

Among the developed techniques, the speaker recognition system widely uses short-term features for speech analysis; MFCC and its derivatives as robust feature extraction techniques. The combination of various variant of MFCC and

the other feature extraction techniques can be considered according to the intended application. In summary, score level fusion of short term features, mainly LPCC & MFCC, based speaker identification system performs well as compared to other short term feature extraction methods and fusion levels.

ACKNOWLEDGEMENT

The author acknowledge the help provided in carrying out this study by Dr. M. S. Vijayaraghawan, Air(Cmdr) V. Sehgal, Shri Manuj Modi and the author is also grateful to Chairman, NTRO for providing infrastructural support, and necessary permission for the study.

REFERENCES

- [1] J. Campbell, "Speaker recognition: a tutorial", *Proceedings of the IEEE*, vol.85, issue 9, pp. 1437–1462, Sep 1997.
- [2] Tomi Kinnunen, Haizhou Lib, "An Overview of Text-Independent Speaker Recognition: from Features to Supervectors", *Speech Communication*, vol. 52, Issue 1, pp. 12–40, Jan 2010.
- [3] H. A. Murthy, F. Beaufays, L. P. Heck and M. Weintraub, "Robust Text-Independent Speaker identification over Telephone Channels", *IEEE Transaction on Speech and Audio Processing*, vol. 7, No. 5, Sep 1990.
- [4] H. S. Jayanna, S. R. Mahadeva Prasanna, "Analysis, Feature Extraction, Modelling and Testing Techniques for Speaker Recognition", *IETE Technical Review*, vol. 3, Issue 3, May-June 2009.
- [5] Q. Zhu, A. Alwan, "On the use of variable Frame Rate Analysis in Speech recognition", *International Conf. on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1783-1786, 2000.
- [6] K. M. Ponting, S. M. Peeling, "Experiments in Variable Frame Rate Analysis for Speech Recognition", Defense Technical Information Centre OAI-PMH Repository (United States), Dec 1989.
- [7] D. Pekar, N. Jakovljevic, M. Janev, D. Miskovic, V. Delic, "On the use of Higher Frame rate in the Training Phase of ASR", ACM Digital Library, *Pro. of the 14th WSEAS International conf. on Computers*, vol. 1, pp. 127-130, 2010.
- [8] Sam Kwong, Qian-Hua He, "The Use of Adaptive Frame for Speech Recognition", *Journal on Applied Signal Processing*, vol 2, pp. 82–88, 2001.
- [9] H. S. Jayanna, S. R. Mahadeva Prasanna, "Limited data speaker identification", *Indian Academy of Sciences*, vol. 35, Part 5, pp. 525–546, Oct 2010.
- [10] Chi-Sang Jung, Moo Young Kim, Hong-Goo Kang, "Selecting Feature Frames for Automatic Speaker Recognition Using Mutual Information", *IEEE Transactions on Audio Speech and Language Processing*, vol. 18, Issue 6, pp. 1332 – 1340, Aug. 2010.
- [11] R. Pawar, H. Kulkarni, "Analysis of FFSR, VFSR, MFSR Techniques for feature Extraction in Speaker Recognition: A Review", *International Journal of Computer Science*, vol. 7, Issue 4, No 1, July 2010.
- [12] A. Oppenheim, R. Schafer, J. Buck, "*Discrete-Time Signal Processing*", second ed. Prentice Hall, 1999.
- [13] J. E. Luck, "Automatic Speaker verification using Cepstral Measurements", *Journal of the Acoustical Society of America*, Vol. 46, Issue 4B, pp. 1026-1032, Nov 1969.
- [14] S. Pruzansky, "Pattern-Matching Procedure for Automatic Talker Recognition", *Journal of the Acoustical Society of America*, vol. 35, Issue 3, pp. 354-358, 1963.
- [15] S. Pruzansky, M. V. Mathews, P. B. Britner, "Talker-Recognition Procedure Based on Analysis of Variance", *Journal of the Acoustical Society of America*, Vol. 35, Issue 11, pp. 1877-1877, 1963.
- [16] J. W. Glenn, N. Kleiner, "Speaker Identification Based on Nasal Phonation", *Journal of the Acoustical Society of America*, vol. 43, Issue 2, pp. 368-372, 1968.
- [17] B. S. Atal, "Automatic Speaker Recognition Based on Pitch Contours", *Journal of the Acoustical Society of America*, vol. 45, Issue 1, pp. 309-309, 1969.
- [18] J. J. Wolf, "Efficient Acoustic Parameters for Speaker Recognition", *Journal of the Acoustical Society of America*, vol 51, Issue 6B, pp. 2044-2056, 1972.
- [19] B. S. Atal, "Effectiveness of Linear Prediction characteristics of the speech wave for automatic speaker identification and verification", *Journal of the Acoustical Society of America*, vol. 55, pp. 1304-1312, 1974.
- [20] A. Rosenberg, M. Sambur, "New techniques for automatic speaker verification", *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 23, Issue. 2, Apr 1975.
- [21] M. Sambur, "Selection of acoustic features for speaker identification", *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 23, Issue. 2, Apr 1975.
- [22] J. Markel, B. Oshika, A. Gray, "Long-term feature averaging for speaker Recognition", *IEEE Trans. on Acoustic Speech and Signal Processing*, Vol. 25, Issue 4, pp. 330 - 337, Aug 1977.
- [23] F. Soong, B. Juang, "Line spectrum pair (LSP) and speech data compression", *IEEE International Conf. on Acoustics Speech and Signal Processing*, pp 37–40, Mar 1984.
- [24] D. Chow, W. Abdulla, "Robust Speaker Identification Based on Perceptual Log Area Ratio and Gaussian Mixture Models", *8th International Conf. on Spoken Language Processing*, vol. III, pp. 1761-1764, 2004.
- [25] F. zohra Chelali, A. Djeradi, R. Djerad, "Speaker Identification System based on PLP Coefficients and Artificial Neural Network", *Proceedings of the World Congress on Engineering*, London, U.K., vol II , July 2011.

- [26] P. Mermelstein, "Distance Measures for Speech Measurements – Psychological and Instrumental", *Joint workshop on Pattern Recognition and Artificial Intelligence*, Hyannis, Mass, June 1976.
- [27] S. Davis, P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences", *IEEE Transactions on Acoustics Speech and Signal Processing*, vol 28 Issue 4, pp. 357 – 366, Aug 1980.
- [28] D. A. Reynolds, R. C. Rose, "Robust Text-Independent Speaker Identification using Gaussian Mixture speaker model", *IEEE Transaction on speech and Audio processing*, vol. 3, No. 1, Jan 1995.
- [29] M. R. Hasan, M. Jamil, MGRMS. Rahman, "Speaker Identification using Mel frequency Cepstral Coefficients", *3rd International Conf. on Electrical & Computer Engg*, Dhaka, Bangladesh, pp. 28-30, Dec 2004.
- [30] F. Zheng, G. Zhang, Z. Song, "Comparison on Different implementations of MFCC", *Journal of Computer Science & Technology*, vol.16, issue 6, pp. 582-589, Sept. 2001.
- [31] E. Ambikairajah, "Emerging features for speaker recognition", *6th International IEEE Conf. on Information, Communications & Signal Processing*, Singapore, pp. 1–7, Dec 2007.
- [32] Qi Li, Y. Huang, "Robust Speaker identification Using an Auditory Based Feature", *IEEE Int. Conf. on Acoustics Speech and Signal Processing*, pp. 4514-4517, Mar 2010.
- [33] A. Mousa, "Mare Text Independent Speaker Identification based on K-mean Algorithm", *International Journal on Electrical Engineering and Informatics*, vol 3, No. 1, pp. 100-108, 2011.
- [34] A. Lawson, P. Vabishchevich, M. Huggins, P. Ardis, B. Battles, A. Stauffer, "Survey and evaluation of acoustic features for speaker recognition", *IEEE International Conf. on Acoustics Speech and Signal Processing*, pp. 5444-5447, 2011.
- [35] S. Furui, "Comparison of speaker recognition methods using statistical features and dynamic features", *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 29, Issue 3, pp. 342 – 350, Jun 1981.
- [36] G. R. Doddington, "Speaker Recognition—Identifying people by their Voices", *Proceedings of IEEE*, vol 73, pp. 1651-1664, Nov 1985.
- [37] M. Sahidullah, S. Chakroborty, G. Saha, "Improving Performance of Speaker Identification System Using Complementary Information Fusion", *Proceedings of 17th International Conference on Advanced Computing and Communications*, pp. 182-187, 2009.
- [38] S. Chakroborty, G. Saha, "Improved Text-Independent Speaker Identification using Fused MFCC & IMFCC Feature Sets based on Gaussian Filter", *International Journal of Signal Processing*, vol 5, No. 1, pp 11-19, 2009.
- [39] S. Singh, E. G. Rajan, "Vector Quantization approach for Speaker Recognition using MFCC and Inverted MFCC", *International Journal of Computer Applications*, vol 17, No. 1, pp. 0975- 8887, March 2011.
- [40] N. Zheng, T. Lee, N. Wang, P. C. Ching, "Integrating Complementary Features from Vocal Source and Vocal Tract for Speaker Identification" *Computational Linguistics and Chinese Language Processing*, Vol. 12, No. 3, pp. 273-290, Sep 2007.
- [41] M. Grimm, K. Roschel "Robust Speech Recognition and Understanding", InTech , Austria, June 2007.
- [42] J. Ming, T. J. Harzen, J. R. Glass, D. A. Reynolds, "Robust Speaker recognition in Noisy Conditions", *IEEE Transactions on Audio speech and language processing*, vol. 15, No. 5, July 2007.
- [43] D. S. Kim, S. Y. Lee, R. M. Kil, "Auditory Processing of speech signals for robust speech recognition in real world noisy environments", *IEEE Trans. Speech Audio Process*, vol. 7, no. 1, pp 55-69, Jan. 1999.
- [44] Y. Gong, "Speech recognition in noisy environments: A survey", *Speech Communication*, vol. 16, Issue 3, April 1995, pp. 261-291, Dec 1999.
- [45] B. J. Shannon, K. K. Paliwal, "Feature extraction from higher-lag autocorrelation coefficients for robust speech recognition", *Speech Communication*, vol. 48, pp. 1458–1485, Aug 2006.
- [46] A. Dev, B. Parmanand, "A Novel Feature Extraction Technique for Speaker Identification", *International Journal of Computer Applications*, vol. 16, No.6, pp. 0975–8887, Feb 2011.
- [47] T. E. Bachir, A. Benabbou, M. Harti, "Design of an automatic Speaker Recognition System based on Adapted MFCC and GMM methods for Arabic speech", *International Journal of computer Science and network Security*, vol. 10, No. 1, Jan 2010.
- [48] S. Kim, M. Ji, H. Kim, "Robust speaker recognition based on filtering in autocorrelation domain and sub-band feature recombination", *Pattern Recognition Letters*, vol. 3, pp. 593–599, 2010.
- [49] P. Bansal, A. De, S. B. Jain, "Robust Feature Vector Set Using Higher Order Autocorrelation Coefficients", *International Journal of Cognitive Informatics and Natural Intelligence*, vol. 4, issue 4, pp. 37-46, October-December 2010.
- [50] R. Schluter, L. Bezrukov, H. Wagner, H. Ney, "Gammatone Features and Feature Combination for Large Vocabulary Speech Recognition", *IEEE International Conference on Acoustics Speech and Signal Processing*, pp. IV-649 - IV-652, June 2007.
- [51] D. S. Kim, J. H. Jeong, J. W. Kim, S. Y. Lee, "Feature extraction based on zero-crossings with peak amplitudes for robust speech recognition in noisy environments", *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 61 – 64, vol. 1, 7-10 May 1996.

- [52] D. L. Hall, J. Llinas, "Handbook on Multi-sensor Data Fusion", CRC press, Edition 1, June 2001.
- [53] F. Huenupan, N. B. Yoma, C. M., C. Garretton, "Confidence based multiple classifier fusion in speaker verification", *Journal Pattern Recognition Letters, ACM Digital Library*, vol 29 Issue 7, pp. 957-966, May 2008.
- [54] J. Keshet, S. Bengio, "Automatic Speech and Speaker Recognition: Large margin and Kernel methods", John Wiley & Sons Ltd, 2009.
- [55] A. Ross, A. Jain, "Information fusion in biometrics", *Elsevier Science Pattern Recognition Letters*, vol. 24, pp. 2115–2125, 2003.
- [56] J. Kittler, M. Hatef, R. Duin, and J. Matas. "On combining classifiers", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 20, issue 3, pp. 226–239, March 1998.
- [57] J. Kittler, "Combining Classifiers: A Theoretical Framework", *Pattern Analysis & Applied. Springer-Verlag London Limited*, Issue 1, pp.18-27, 1998.
- [58] J. Kittler, F.M. Alkoot, "Sum Versus Vote Fusion in Multiple Classifier Systems", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25 Issue: 1, pp. 110 – 115, Jan 2003.
- [59] K. Chen, L. Wang, H. Chi, "Methods of Combining multiple Classifiers with Different Features and Their Applications to Text-Independent Speaker Identification", *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 11, issue 3, pp. 417-445, 1997.
- [60] K. R. Farrell, R. J. Mammone, K. T. Assaleh, "Speaker Recognition using Neural Networks and conventional classifiers", *IEEE transaction on speech and audio processing*, Vol. 2, No. 1, Part II, Jan 1994.
- [61] R. Tong, B. Ma, K. A. Lee, C. You, D. Zhu, T. Kinnunen, H. Sun, M. Dong, E. S. Chng, H. Li, "The IIR NIST 2006 Speaker Recognition System: Fusion of Acoustic and Tokenization Features" *Proceedings of the 5th International conf. on Chinese Spoken Language Processing, ACM digital Library*, pp. 566-577, 2006.
- [62] P. Thevenaz, H. Hugli, "Usefulness of the LPC-residue in text-independent speaker verification", *Speech Communication archive*, vol. 17, Issue 1-2 Aug 1995.
- [63] S. R. M. Prasanna, C.S. Gupta, B. Yegnanarayana, "Extraction of speaker-specific excitation information from linear prediction residual of speech", *Speech Communication*, vol. 48, pp. 1243-61, 2006.
- [64] M. D. Plumpe, T.F. Quatieri, D.A. Reynolds, "Modeling of the glottal flow derivative waveform with application to speaker identification", *IEEE Trans. Speech Audio Process.*, vol. 7, issue 5, pp. 569-85, 1999.
- [65] B. Peskin, J. Navratil, J. Abramson, D. Jones, D. Klusacek, D. A. Reynolds, B. Xiang, "Using prosodic and conservational feature for high-performance speaker recognition", *Int. Conf Acoustic, Speech, Signal Processing, Hong Kong*, vol. IV, pp. 784-7, Apr. 2003.
- [66] J. P. Campbell, D. A. Reynolds, R. B. Dunn, "Fusing High- and Low-Level Features for Speaker Recognition", *Proc. European Conf. Speech Communication Technology*, pp.2665-2668, 2003.
- [67] Y. A. Solewicz, M. Koppel, "Using Post-Classifiers to Enhance Fusion of Low and High-Level Speaker Recognition", *IEEE Trans. Audio Speech and Language Processing*, Vol. 15, No. 7, Sep 2007
- [68] M. Salkhordeh, H. A. Vahedian, H. Sadoghi, Y. H. Modaghegh, "Designing Kernel Scheme for Classifiers Fusion", *International Journal of Computer Science and Information Security*, vol.6, No. 2, 2009.
- [69] N. M. Wanas, R. A. Dara, M. S. Kamel, "Adaptive fusion and co-operative training for classifier ensembles", *Elsevier Pattern Recognition*, vol 39, pp.1781 – 1794, 2006.
- [70] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene, A. Gelzinis, "Soft combination of neural classifiers: A comparative study", *Elsevier Pattern Recognition Letters*, vol. 20, pp. 429-444, 1999.
- [71] R. Tong, B. Ma, K. A. Lee, C. You, D. Zhu, T. Kinnunen, H. Sun, M. Dong, E. S. Chng, H. Li, "Fusion of Acoustic and Tokenization Features for Speaker Recognition", *Chinese Spoken Language Processing Lecture Notes in Computer Science*, vol 4274, pp 566-577, 2006.
- [72] B. Reuderink, M. Poel, K. Truong, R. Poppe, M. Pantic, "Decision-Level Fusion for Audio-Visual Laughter Detection", *Springer-Verlag Berlin Heidelberg*, pp. 137–148, 2008.
- [73] A. Mishra, "Multimodal Biometric it is: Need for Future Systems", *International Journal of Computer Applications*, vol. 3, No. 4, pp. 28- 33, June 2010.
- [74] Z. Wu, L. Cai, H. Meng, "Multi-level Fusion of Audio and Visual Features for Speaker Identification" *Springer-Verlag Berlin Heidelberg*, pp. 493– 499, 2005.
- [75] N. Brummer, L. Burget, J. Cernocky, O. Glembek, F. Grezl, M. Karafiat, D. Leeuwen, P. Matejka, P. Schwartz, A. Strasheim, "Fusion of heterogeneous speaker recognition systems in the STBU submission for the NIST speaker recognition evaluation 2006", *IEEE Trans. Audio Speech and Language Processing* vol.15, No.7, pp. 2072–2084, Sep 2007.
- [76] B. Reuderink, M. Poel, K. Truong, R. Poppe, M. Pantic, "Decision-Level Fusion for Audio-Visual Laughter Detection" *Machine Learning for Multimodal Interaction Lecture Notes in Computer Science Springer-Verlag Berlin Heidelberg*, vol 5237, pp. 137–148, 2008.
- [77] T. Kinnunen, V. Hautamäki, P. Fränti, "Fusion of Spectral Feature Sets for Accurate Speaker Identification", 9th Conference Speech and Computer, Saint-Petersburg, Russia Sep 20-22, 2004
- [78] L. Valet, G. Mauris, P. Bolon, "A Statistical Overview of Recent Literature in Information Fusion" *IEEE*

- Magazine on Aerospace and Electronic Systems, vol. 16, Issue: 3, pp. 7 – 14, 2001.
- [79] M. R. Islam, M. F. Rahman, “Hybrid Feature and Decision Fusion based Audio-Visual Speaker Identification in Challenging Environment”, *International Journal of Computer Applications*, vol. 9, No.5, Nov 2010.

Innovative Use of Cloud Computing in Smart Phone Technology

Vaibhav Rana

Submitted in March, 2013; Accepted in August, 2013

Abstract - This paper elaborates the vital and innovative use of Cloud computing in the field of Smart Phone Technology. In the proposed environment there is a cloud server installed at the central place called as cloud center of mobile computing. The function of cloud computing is to accept the request of the end users and process the required applications accordingly. Henceforth this server is fully equipped with all advance ICT infrastructure, mobile communication utilities and full security measure, so that every feature can be used with full security. All users are having equal priority to get connection with the cloud server and use every application as well as each function equally. Due to use of this modern technology the main advantage is that a single entity is being used by many remote users for various requirements at the same time without occupying much space and infrastructure. The resulting new paradigm of mobile cloud computing is being embraced by researchers, professionals and practitioners as an exciting new way to extend the capabilities of mobile devices and mobile platforms. This has the potential for profound impacts on the business environment and people's daily life.

The most salient feature of the system is that by the use of this infrastructure we can perform mobile communication and electronic communication by a single unit and the same communication spectrum. The presented cloud computing technology offers a natural solution for expending the limited capabilities of mobile devices. This process makes the whole communication system cost effective, efficient, transparent, modular system.

Index Terms - MTSO, IaaS, SaaS, IaaS, NaaS, Virtual Machines Cloud Client, Base Station, Cell Site, PSTN, Mobile Station.

1. INTRODUCTION

Recent advanced researches in 'Mobile Communication Networks' and increasing penetration of smart phones are transforming the 'Mobile Internet' and empowering end users with rich mobile experience. Although having the limited computing utility features, and storage capabilities of mobile devices are affecting the customer mentality, this hampering their abilities to support increasingly demand of sophisticated applications by the end users. As described in [1,2], we know

that communication is a very important process in human's life, so that human being continuously evolving this process day by day and making it simpler and faster. Cloud computing is one of the very latest step in this direction. Cloud computing is the use of computing resources (hardware, software and networking) that are delivered as a service over a network (typically the Internet). The name comes from the use of a cloud-shaped symbol as an abstraction for the complex infrastructure. It represented in system diagram fig.1. Cloud computing entrusts remote services with a user's data, software and computation. Thus it is just like using some applications or facilities by not directly installing in devices as we normally do. This system is the evolved version of remote access. The figure below shows that all applications, infrastructures are present at a centralized location, from there all end users can access these applications and infrastructures.

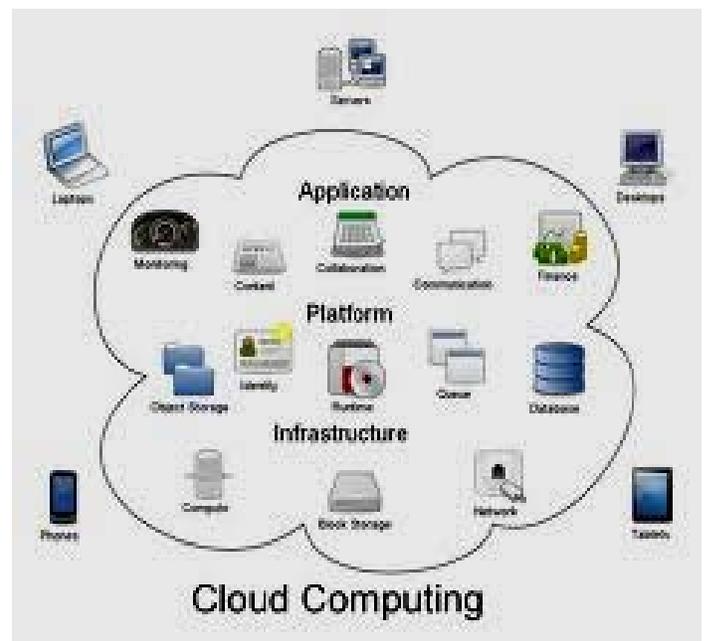


Figure 1: Cloud Computing Architecture

Cellular Communication is one of the old and still growing communicating ways in today's world. Birth of this process is due to Marconi who invented antenna and successfully perform the experiment of transmitting and receiving the electromagnetic wave. From that day wireless communication can't be stopped in communication movement. To make efficient use of antenna in communication many ways are used and finally cellular communication comes into picture. Now a day's cellular communication has become a necessity for communication architecture. As conceptualized in [3], cellular

Department of Electronics & Communication, Rajiv Gandhi
Proudyogiki Vishwavidyalaya, Bhopal
E-mail: ranavaibhav@hotmail.com

system or mobile network is a radio network distributed over land areas called as cells. Each cell is served by at least one fixed-location transceiver, known as a cell site or base station. In a cellular network, every cell uses a different set of frequencies from neighboring cells. It happens to avoid interference and provide guaranteed bandwidth within each cell. When combined together these cells provide radio coverage over a wide geographic area. This enables a large number of portable transceivers (e.g., mobile phones, pagers, etc.) to communicate with each other using fixed transceivers and telephones anywhere in the network, via base stations. Even if some of the transceivers are moving through more than one cell during transmission. Cellular networks offer a number of advantages over alternative solutions as:

- Flexible enough to use the features and functions of almost all public and private networks.
- Increased capacity reduced power use.
- Larger coverage area.
- Reduced interference from other signals.

The most common example of a cellular network is a mobile phone (cell phone) network stated in [3]. A mobile phone is a portable telephone which receives or makes calls through a cell site (base station), or transmitting tower. Radio waves are used to transfer signals to and from the cell phone. Modern mobile phone networks use cells because radio frequencies are a limited, shared resource. Cell-sites and handsets change frequency under computer control using low power transmitters, so that a limited number of radio frequencies can be used by many callers simultaneously with less interference. A cellular network is used by the mobile phone operator to achieve both coverage and capacity for their subscribers. Large geographic areas are split into smaller cells to avoid line-of-sight signal loss and to support a large number of active phones in that area. All of the cell sites are connected with telephone exchanges (or switches), which in turn connect to the public telephone network in cities, each cell site may have a range of up to approximately ½ mile, while in rural areas, the range could be as much as 5 miles. It is possible that in clear open areas, a user may receive signals from a cell site 25 miles away. Since almost all mobile phones use cellular technology, this includes GSM, CDMA, and AMPS (analog). The term "**cell phone**" is in some regions, notably the US, used interchangeably with "**mobile phone**". However, **satellite phones** are mobile phones that do not communicate directly with a ground-based cellular tower, but may do so indirectly by way of a satellite as shown in **fig.2**.

1.1 CLOUD CONCEPTION

Cloud computing is a growing computing paradigm that was innovated to deploy cost effective solutions over Internet. As mentioned in [4], various companies such as Google, IBM, Amazon, Yahoo and Intel have already started providing computing infrastructures for its intended use. Cloud computing is a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet. It is a scalable service delivery

platform build over Service Oriented Architecture (SOA) and virtualization concepts. The benefit of cloud can be stated in two perspectives as:

- From cloud service provider perspective and
- From cloud user perspective.

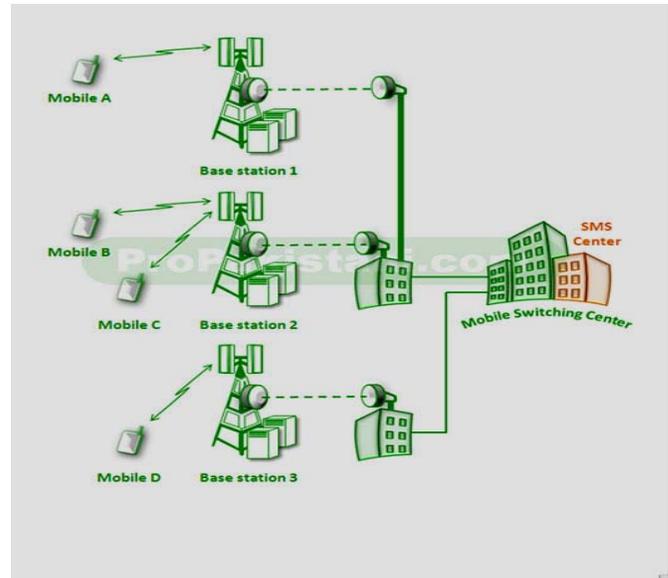


Figure 2

Cloud service providers get benefited with the better utilization of infrastructure they own. Even, they can obtain an improvement in the processing power with minimal additional cost as defined in [7, 8, 9, and 10]. The article [13] have described the various burning issues related to mobile application with cloud computing. Thus, giving way for new business opportunities and speed up existing business processes. Users get benefited by paying only for what they consume and even they need not have insights of the technology infrastructure used in the “cloud” that supports them. Data analytics applications which require processing extremely large datasets while supporting extreme scalability is attracting huge interest towards cloud. The cloud computing generally incorporates combinations of the following:

- Providing servers, software, and office space and network equipment as service.
- Providing computing platform and solution stack as a service.
- Providing an application as a service.

Cloud contains of large number of servers facilitating distributed computing platform. Each server can have multiple virtual machine instances (VMs) of the applications hosted in the cloud. As customer demand for those applications changes, new servers are added to the cloud or idled and new VMs are instantiated or terminated. A major advantage with cloud is the recoverability in case of disaster. Cloud data is mostly backed up regularly and stored in a different geographic location. This eases the recovery process which otherwise is a very costly feature for companies. Cloud computing infrastructure is totally

different from the current data warehouse infrastructures. Unlike the high-end servers which constitute the current infrastructure, Cloud computing offers the same functionality at low-cost. Since most analytics for an organization use large number of database reads than writes. There is a requirement for new relational database software architecture to efficiently store and retrieve large volumes of data for computation on Cloud. In a cloud, node failures and node changes may occur frequently. Hence the cloud uses a vast numbers of processing units, so the failures may occur in the cloud also. It can be shielded from its users using robust policies to make the cloud as tolerant as required (or requested). The best cloud databases will replicate data automatically across the nodes in the cloud cluster. This would be able to continue running in the event of 1 or more node failures, and be capable of restoring data on recovered nodes automatically—without DBA assistance. Ideally, replication will be “active-active” in that the redundant data may be queried to increase performance. The whole cloud computing is mainly divided into 3 steps as the 3 cloud services are given below:

- a) Infrastructure as a service
- b) Platform as a service
- c) Software as a service

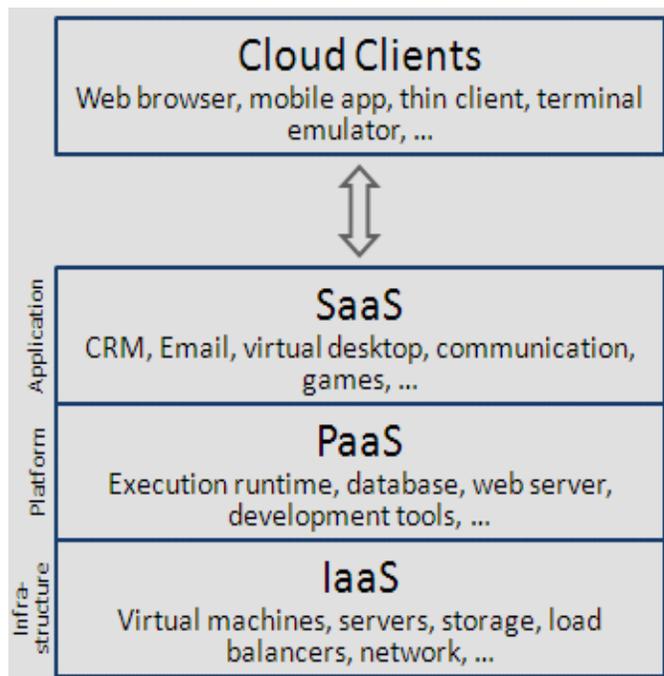


Figure 3

1.1.1 Infrastructure as a Service

It is the most basic service provided by service provider. It broadly contains database, computing server, network, etc. In the most basic cloud-service model, providers of IaaS offer computers - physical or virtual machines - and other resources. Clouds often offer additional resources such as images in a virtual-machine image-library, raw (block) and file-based

storage, firewalls, load balancers, IP addresses, virtual local area networks (VLANs), and software bundles. To deploy their applications, cloud users install operating-system images and their application software on the cloud infrastructure. In this model, the cloud users install and maintain the operating systems and the application software. So by the above explanation we can say that in first step we are developing basic requirement for cloud computing the machines, network, storage, operating system, etc. The service provider will arrange all the basic requirements on a single place and control the whole communication process.

1.1.1. Platform as a Service

Now the next step is an intelligibility step because here the service provider has to use the infrastructure which was collected in first step and is used to make the software run on that infrastructure, so that the desired results can be obtained. In the PaaS model, cloud providers deliver a computing platform typically including operating system, programming language execution environment, database, and web server. Application developers can develop and run their software solutions on a cloud platform without the cost and complexity of buying and managing the underlying hardware and software layers.

1.1.2 Software as a Service

By using above two services the service provider can put various applications in the cloud server which are used by end user. In the SaaS model, cloud providers install and operate application software in the cloud and cloud users access the software from cloud clients. The cloud users do not manage the cloud infrastructure and platform on which the application is running. This eliminates the need to install and run the application on the cloud user's own computers simplifying maintenance and support. What makes a cloud application different from other applications is its scalability. This can be achieved by cloning tasks onto multiple virtual machines at run-time to meet the changing work demand. Load balancers distribute the work over the set of virtual machines. This process is transparent to the cloud user who sees only a single access point. To accommodate a large number of cloud users, cloud applications can be multitenant, that is, any machine serves more than one cloud user organization. So there are many services in the form of software are kept on the cloud network which is used by end user.

1.1.3 Network as a Service

As the name showing, provider uses network to provide the services to the end user, inter network of cloud network. A category of cloud services where the capability provided to the cloud service user is to use network/transport connectivity services and/or inter-cloud network connectivity services. Traditional NaaS services include flexible and extended VPN, and bandwidth on demand. NaaS concept materialization also includes the provision of virtual network service by the owners of the network infrastructure to a third party. Now some special

arrangement must be done like according to the requirement of user, that user must have some minimum hardware and software.

1.1.4 Cloud Client

In cloud computation end user is called cloud client. So the cloud client must fulfill some minimum requirement. User access cloud computing using networked client devices, such as smartphones. Some of these devices (cloud client) rely on cloud computing for all or a majority of their applications so as to be essentially useless without it. Examples are thin clients and the browser-based clients. Many cloud applications do not require specific software on the client and instead use a web browser to interact with the cloud application. Some legacy applications (line of business applications that until now have been prevalent in thin client Windows computing) are delivered via a screen-sharing technology as studied from [6]. Finally, on the basis of all above explanation we come to know that this whole arrangement is centralized in providing the services and control the communication.

1.2 Cellular Mobile Communication

There are many reasons for the development of mobile communication over conventional telephone system. These are limited service capability, poor service performance and inefficient frequency spectrum utilization, etc.

1.2.1 Basic Cellular System

A basic cellular system consists of three parts: a mobile unit, a cell site and a mobile telephone switching office (MTSO) is presented in fig.4.

A mobile unit contains a control unit, a transceiver and an antenna system. Mobile unit is the basic unit of cellular mobile communication. It is operated by end users. This is the device by which the end user can avail the facilities and communicate to other end user. As the technology is growing this basic unit has also evolved too much. Now a day's smartphone has replaced the most basic mobile phone. A mobile phone contains a central processing unit which performs the most basic function of communication i.e. modulation and demodulation. This unit contains a transceiver system to perform modulation and demodulation for transmitting and receiving the data/information. Finally, a mobile unit must contain an antenna to radiate and receive the radio waves which contains the information transmitted in wireless communication process. Here in this study mobile unit is smartphone. Smartphone is a mobile unit with a dedicated operating system with very advanced computing abilities, advance connection capabilities than a normal phone. Smartphone is a combination of PDA (personal digital assistant), normal mobile phone, functionality of portable media players, low end compact digital cameras, pocket video cameras, and GPS navigation units to form one multi-use device.

1.2.3 Cell Site

The cell site provides a interface between the MTSO and the mobile units. It has a control unit, radio cabinets, antennas, a power system and data terminals. Cell site is a mediator because of cell site only a connection can be established. This contains a central processing unit which is responsible for making the connection of a mobile unit with MTSO; it is also responsible for basic requirement of communication. This CPU have to do many tasks like establishing connection, handoff, maintaining the signal strength, etc. The next part of cell site is radio cabinets. These are used to control the radio waves, which are finally used in communication. The next part is antenna, it is very important part of a cell site to maintain proper signal strength and low noise communication then antennas plays a very big role in these areas, so selection of proper antenna is required. This whole system is named as base station. Over this the designing of cell site is also very important to achieve the good communication process. The designing also plays a very important role to save bandwidth, spectrum, power, etc. As drawn in fig.5.

1.2.4 Mobile Telephone Switching Office(MTSO)

The switching office, the central coordinating element for all cell sites, contains the cellular processor and cellular switch. It interfaces with telephone company zone offices; controls call processing and handle billing activities. It is the main soul of whole cellular communication process because it switches the connection between one end user to another end user by making connection with another MTSO through wire or satellite. It records billing, call records, data usages and all information of an end user like registration id, address, etc., as shown in [5].

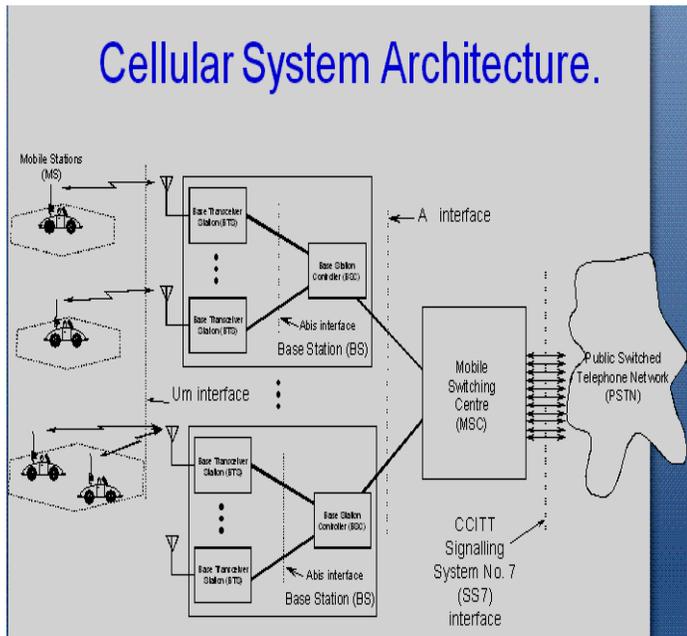


Figure 4

1.2.2 Mobile Units

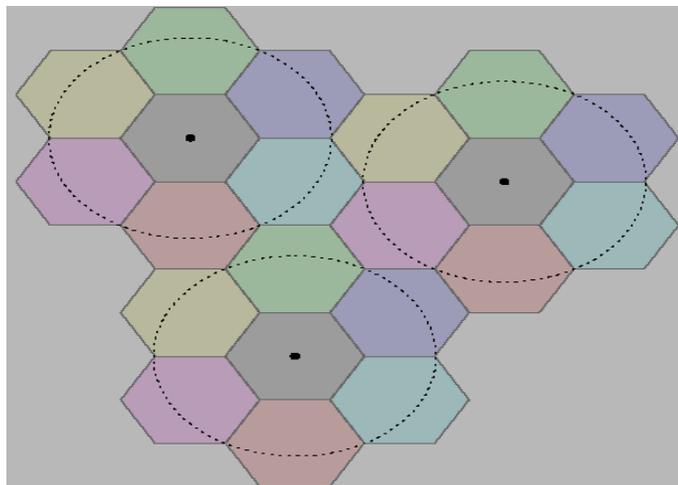


Figure 5

2. SMART PHONE TECHNOLOGY

Actually a smart phone is nothing but a mobile phone built on a mobile operating system with more advanced computing capability connectivity than a feature phone. The first smart phone combines the functions of a personal digital assistant (PDA) with a mobile phone. Now a days many modern smart phones also includes high resolution touch screens and web browsers that display standard web pages as well as mobile optimized sites. High speed data access is provided by Wi-Fi and mobile broad band. In recent years the rapid development of mobile application market has been drivers of smart phone adoption. One of the significant differences is that the advanced application programming interfaces (APIs) on smart phones for running third party applications can allow those applications to have better integration with the phone's OS and hardware than is typical with features phones. A Smart Phone is one device that can take care of all required hand held computing and communication needs in a single small box. This study explores what makes a smart phone and what we can do with it. As cell phones and PDAs are the most common handheld device now, a smart phone is usually either a phone with added PDA capabilities or a PDA with added phones capabilities. A list of things which a smart phone can do is as:

- 1) Send and receive mobile phone calls some smart phones are also Wi-Fi capable.
- 2) Personal information (PIM) including notes, calendar and to do list.
- 3) Communication with Laptop or desktop computers.
- 4) Data synchronization with applications like Microsoft outlook and Apple's ical calendar programs
- 5) E-mail
- 6) Instant messaging
- 7) Application such as word processing or video games

Smart phones have recently gained significant popularity in heavy mobile processing while users are increasing their expectations towards rich computing experience. However resource limitations and current mobile computing advancements hinder this vision.

3. CLOUD COMPUTING

Everything may be hosted in the clouds, it is a large group of interconnected computers. Users need not have knowledge of expertise in or control over the technology in cloud. Various scripts [14, 15, and 16] advocates the generalization of automatic cloud computing and cloud engine architecture. Let us see how cloud computing works with a smart phone. You can use the entire capacity of the hand device for music or apps, because you can store your documents online and access them. No matter where you are you can even buy an application for your business and share it with your colleagues by using one of these services, but your client has to limit himself/herself to smart phones only. You can take advantage of all of these things from home by using your tablet, since it is running on the same Operating System as your smart phone. Technology like this will give users more device choices while simplifying its provisioning duties and maintaining security of corporate data assets in place of physical phones. It would manage and provision virtual machine profiles that execute on phones purchased by employee. Now a days user are making such heavy use of cloud based services that local storage has become something of an afterthought. As a result, we have less data to store on our phones. We look at the mobile device features with storage as secondary concern, so vendor could save money on storage costs and invest it on other hardware features. As a result traditional bottlenecks could be alleviated and the overall hand set experience could be improved.

4. WORKING SYSTEM METHODOLOGY

Script [12] expressed a new concept as NEC enables smartphone access to work environment on the cloud as secure access to office data from smartphones. In this mechanism the employee access their work while away from the work station by virtualization of smart device such as smart phone and tablets.

In past few years smart devices have become broadly utilized through the world and an unavoidable need has raised by employee to easily access work information from their smartphone when far away from the office. Henceforth introduced a cloud based smart phone access to work station. By using this facility the NEC cloud Smart phone users can safely and securely access office data from remote locations with these smart phones, since there is no risk of leaking information from a lost device. By viewing this fact I have applied the similar approach in the field of cloud oriented smartphone process.

As an end user switch on the mobile station, the initialization work starts. The frequencies which are used here are mainly 450MHz, 950MHz, 1800MHz, 2100MHz. Mobile station must tune to the strongest dedicated control channel within few seconds. If it cannot complete this task on the strongest dedicated control channel, it attempts to the second strongest channel and again if this task is not completed then mobile station search for third channel and so on. When the end user starts to access application then an access timer is set for

origination maximum 12 second, page response maximum 6 second, an order response maximum 6 second, registration maximum 6 second. Now access channel is scanned by mobile station and tuned to the strongest channel. When tuning is done then the mobile station is authenticated by the base station. In authentication base station verify the identification of the user by comparing the information already present with base station. After verification, base station acknowledges the mobile station. The mobile station gets the acknowledgement then only user is able to avail the services provided by the service provider. This whole process is too fast that user is not able to notice it. As soon as user accesses the service the mobile station sends the service request to the base station in response to request base station. This acknowledges the mobile station to inform the mobile station that it is able to access that service. Base station will make connection to the MTSO at the time of installation and this connection is maintained forever. The connection of base station and MTSO is mostly wireless through satellites. The main work of base station is to enhance the communication process in mobile communication by correlating the high frequencies used in satellite communication with low frequencies used in normal wireless communication. Now base station will forward the mobile station's request to the MTSO through satellite. After receiving the signal from base station MTSO also acknowledges the base station. Now the request is processed in server of MTSO so that the requested service can be provided. After computing on request the server locates the nearest possible virtual machine so that the requested service can be provided. The reverse way to the end user is from MTSO to base station through satellite communication, base station to mobile station through normal wireless communication. This whole responding process is also very fast while the end users are unable to notice it. This whole process will be same for all users who want to access the service.

In the present study the application of cloud computing is associated with smart phone technology. We have implemented the above two strategies together to form as a unique communication system i.e. mixing of cloud computing and cellular mobile communication. This can be done by combining MTSO and Cloud Server. Because of this, intelligibility of MTSO will be enhanced by the use of IaaS, PaaS, SaaS services of cloud computing. Thus the connection process is too much simplified as well as easier that the service provider gets the full control over whole communication process. Mobile network also help the cloud computing by providing very large established network. To maintain uninterrupted cellular communication the entire network is well planned and designed.

5. RESEARCH ELEMENT OF CLOUD COMPUTING

Cloud computing consists a wider range of controls, technologies and policies used to protect the existing infrastructure, applications and data. Security issues experienced by end user or by cloud suppliers is of prime concern. Henceforth cloud providers must make sure that what

they are offering is secure and customer's applications as well as data are protected. Simultaneously the client on the other hand must ensure that the cloud supplier has the appropriate security measures in order to protect his data and applications. Generally cloud computing elements come into four categories:

1. **Legal Issues:** End users and cloud vendors have to negotiate about liability, end-of-service and both must agree about the degree of liability of each party.
2. **Compliance:** Cloud computing providers must be able to provide audit trails and logs. System must be maintained, secured properly and be accessible in case of forensic investigation takes place.
3. **Privacy and security:** Cloud data center must be maintained in such a way that they adhere to compliance requirements in term of privacy.
4. **Identity Managements:** Every user must have his identity management system in order to access computing and information resources. Cloud providers must able to provide such system to their users. They should ensure that users can easily access their application and data as and when they require.

On the basis of latest research work there are several elements of cloud computing based on smart phone. Out of those the value types are as follows:

- **Economic Element:** Economic value of cloud is an alignment of time and size of the investments. You make with the value you receive referred to as 'pay as you go' and 'pay as you grow'.
- **Architectural Element:** The architectural value of cloud is associated with consistent abstract environment presented to developers and operations folks that hides a lot of complexity, making it much simpler and quicker to develop and deploy applications.
- **The Strategic Element:** The strategic value of cloud is a platform which helps you to focus on what makes your organization more effective and different.
- **Type of Cloud Service:** Development PaaS-Programming environment IaaS. Servers Software and Spaces Web Services Utility Computing, Software as service Access to an application via API.
- **Barriers Technical Issues:** Business Model issue, Internet issue, Security issue Compatibility issue ,Social issue.
- **Who should not Use:** The Internet-Impaired, Offline workers, the security Conscious.
- **Who Benefits from Cloud Computing| Collaborator:** Share your documents, Road warrier, Access it from anywhere, lost conscious, user/IT Department user with increasing needs.
- **Disadvantages:** Requires a contact Internet connection, Does not work well with low- speed connectivity, Stored data might not be secured if the cloud loses your data.
- **Advantage:** Data safety, Improved Compatibility between operating systems, Can connect the windows computer to the cloud and share the documents with computers running Apple's Mac OS, Linux, Unix, Remove the device

dependency, Lower cost Computing for users, Improved performance of lower IT infrastructure, Unlimited storage capacity.

- **Cloud Services:** Web based services offered via cloud computing, Accessed by browser, if PC crashed it does not affect either the application or documents, Access from anywhere.
- **Cloud Storage:** Primary use of clouds, multiple third party services as used rather than on the traditional dedicated servers.
- **Virtual Services:** Appears as if data stored in a particular place with a specific name but that place does not exist in reality. Cheaper than dedicated physical resources, secure from accidental erasure or hardware crashes.

Characteristics of Cloud Computing:

User Centric- Data in the cloud belongs to user and can share.

Task Centric- Focus is shifted from what it can do to what you need done. You can retrieve more information as many parallel computers are running your application

Intelligent- Data mining and analysis are very much in need as a very large set of data stored on the computers in cloud system.

Programmable: Cloud's computers must be under the supervision of IT experts so that tasks performed in cloud system can automate the redistribution and data can be locally shared.

6. SYSTEM SECURITY

Security is a basic concern in the system because end user is using the infrastructure of service provider. Considering the latest challenges of data corruption, data stealing, data threatening all the advance security measures are adopted. These measures are to implement the latest firewall facility, latest version of, antivirus, malware protection features, opening of only required communication ports, updating of latest patches and preventive maintenance and many more, as mentioned in [17]. So that end user trusts on the service provider that his/ her data is safe. By following the said precautions our cloud oriented smartphone computing system is reliable and trust worthy

7. FACILITIES GETTING THROUGH THIS SYSTEM

The facilities that can be obtained by using cloud computing with smartphone technology will be uncountable but same major benefits are:

- Area of communication is increased by combining these different cellular communication as well as Internet communication technologies. It can enhance the cumulative properties of both.
- Video conferencing has become fast and convenient.
- Everyone gets equal opportunity to access all the data and every facility.
- Calling, sms, mms, chat, etc. has come on a single platform and available to all.

- Security features can be applied centrally to the whole network with an ease and those features will be technically advanced.
- Progress in any field will benefit the whole communication system.
- Control and management of whole arrangement will be easy and very fine.

8. SYSTEM OUTCOME

The main motive of enhancing any technology is to get best outcome of those enhancement that change the attitude of society and influence all the sectors of the society like education, health, agriculture, economy, trade and industry and day to day human life requirement. Some enhancements of this modular infrastructure are:

8.1 Social Outcome

Very first outcome that must be checked is social impact, it is very important to study the in reference of society welfare. Now a day's smart phone technology is playing a dynamic role in the society. Some few impacts are given below:

- Common chances to everyone to communicate and access to data bank/data resources.
- Equal rights to everyone in communication and utilization of human resources.
- After the implementation of this system worldwide there is no distance between everyone.
- Communication among the society is fast, transparent and reliable
- Equal right and chance for anyone in the society.
- Detection of crimes/frauds is very easy.
- Useful tool to increase the business related activities.
- Can become a major developing key factor for trade and industry,
- Best media to extend the information and messages
- Cost effective media for communication and transmission of information, data etc.
- Transmitting and receiving the emails.
- Receiving and sending the public welfare related policies, procedures and protocols.

8.2 Educational outcome

As education is a very important and powerful skill in this world and as time passes the level and system of education is also changes now technologies playing a very important role in enhancement of this system. This arrangement can bring revolution in the field of education, so there is an impact over education system. So few impacts are:

- Improvement in distance learning.
- Enhancement in interactive learning.
- Best for tutorial learning.
- Provide access to vast library globally.
- Improvement in practical learning by accessing wide laboratories.
- Best knowledge bank can be formed with ease.

- Education could be provided to everyone without any difficulty.
- Powerful internet learning tools.
- Knowledge strengthen by web surfing.
- Knowledge sharing with international universities and institutions.
- Knowledge networking of GIS worldwide.

8.3 Economical outcome

Now the most important impact is on economy because to grow any society in all respect economy has to also growth in every field. So by this arrangement economic benefits must be considered for the growth of this technology and society. Some impacts on economy are:

- Money can be saved, because when these two technologies are combined, progress in any one field will serve to both and finally budget on communication process will be reduced.
- Infrastructure requirements are low due to which space use is reduced.
- As infrastructure is low so maintenance expenditure is low.

Now a day's communication has become so simple and easy due to cellular mobile communication that even languages can't form any barrier in communication. In today's world smartphone is so common that every second person is having it. So it has become one of the basic necessities due to this only communication is easy and fast.

9. PROPOSED RESEARCH TOOL

Tool which are being used in this study are:

- Google docs, it is used to provide service to end user to create documents, manuscript, presentation, spreadsheets, etc. Which help end user to provide office utilities without any installation of any special office software.
- Mindmeister, it is used to draw mindmaps, schematic diagrams which help in designing networks which can be used in building networks for communication purpose.
- Mendeley, it is an online sharing application which helps in building online research libraries.
- Rational suite with some advancement may be implemented to enhance the cloud server.
- Really simple syndication (RSS) is also a major tool to strengthen the cloud based mobile computing network system.

10. CONCLUSION

Finally, by viewing the contents of the manuscript we can observe that the combination of these two technologies will influence the whole world and after some time it is impossible to differentiate between people on the basis of language. Communication has become much simple, better and fast globally. Huge saved money may be utilized for other developmental activities of the society. Knowledge banks are ready to provide the knowledge data to everyone without barrier easily. Educational problems can be resolved easily

through this configuration form society and we are dreaming a better future, better society, etc. As communication is so easy and efficient then latest technologies is easily available to common man with less infrastructure. Mobile computing is become a basic requirement of human being.

11. FUTURE SCOPE

The work can be further extended by applying the security measures which will be developed by studying the loop holes in the proposed system. One way is to develop the application for this system which must be light and user friendly or develop the more easy and advance ways for communication with the application of mobile cloud computing.

NOTE

The author of this paper is BE in Electronics & Communication from Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal and involved in various research and developmental activities.

REFERENCES

- [1]. V. S. Rana, 'An Innovative Use of Information and Communication Technology (ICT) in Trade Facilitation in India', BVICAM's International Journal of Information Technology, BIJIT, Issue 8: (July - December, 2012 Vol.4 No.2)
- [2]. Avijit Dutta, 'Digital Communication and Knowledge Society India', BVICAM's International Journal of Information Technology, BIJIT, Issue 8: (July - December, 2012 Vol.4 No.2)
- [3]. Ashema Hasti, 'Study of Impact of Mobile Ad hoc Networking and its Future Applications', BVICAM's International Journal of Information Technology, BIJIT, Issue 7: (January-July, 2012 Vol.4 No.1)
- [4]. Rajiv and Manohar Lal, 'Web 3.0 in Education & Research', BVICAM's International Journal of Information Technology, BIJIT, Issue 6: (July-December, 2011 Vol.3 No.2)
- [5]. B. V. Ramanamurthy, K. Srinivas Babu and Mohammed Shafruddin, 'Dynamic Data Updates for Mobile Devices by Using 802.11 Wireless Communications', BVICAM's International Journal of Information Technology, BIJIT, Issue 5: (January-June, 2011 Vol.3 No.1)
- [6]. 'Mobile Cloud Computing' available at: http://en.wikipedia.org/wiki/Mobile_cloud_computing
- [7]. 'Future prospects of Mobile Cloud Computing' discussed at http://www.comsoc.org/files/Publications/Magazines/wc_i/cfp/cfpwcm0613.htm
- [8]. 'Application and benefits of mobile cloud computing' available at http://readwrite.com/2009/08/04/why_cloud_computing_is_the_future_of_mobile
- [9]. 'Cloud computing' presented at http://en.wikipedia.org/wiki/Cloud_computing

- [10]. 'Trends in Mobile Cloud Computing' introduced at <http://www.ibm.com/developerworks/cloud/library/cl-mobilecloudcomputing/>
- [11]. 'Mobile Cloud computing in small business management' discussed at <https://www.brighttalk.com/webcast/1141/29233>
- [12]. 'NEC Enables Smartphone Access to Work Environment on the Cloud' available at <http://www.nec.com/en/press/>
- [13]. 'Mobile Application with cloud computing', International Journal of Scientific and Research Publications, Vol.2, Issue 4, April 2012 ISSN 2250-3153
- [14]. S. Pandey, D. Karunam and R. Buyya, "Workflow Engine for Clouds" Clod Computing: Principles and Paradigm, Wiley, 2011.
- [15]. R. Ranjan, R. Buyya, and M. Prashar, "Automatic Cloud Computing: Technologies, Services and Applications, Concurrency and Computation: Practice and Experience, 24(9):935-937, Wiley Press, New York, USA, June 2012.
- [16]. J.O. Kephart and D.M. Chess, The Vision of Automatic Computing, Computer, 36(1):41-50. IEEE, Jan. 2003.
- [17]. M.H. Sqalli, F. Al-Haidari, and K.Salah. EDoS-Shield- A two-stp mitigation technique against EDoS attacks in Cloud Computing, Processing of the 4yh International Conference on Utility and Cloud Computing (UCC 2011), Melbourne, Australia, 2011

A Novel Hénon Map Based Adaptive PSO for Wavelet Shrinkage Image Denoising

Shruti Gandhi¹, Sonal Doomra², Akash Tayal³ and Ashwini Kumar⁴

Submitted in March, 2013; Accepted in August, 2013

Abstract - Degradation of images due to noise has led to the formulation of various techniques for image restoration. Wavelet shrinkage image denoising being one such technique has been improved over the years by using Particle Swarm Optimization (PSO) and its variants for optimization of the wavelet parameters. However, the use of PSO has been rendered ineffective due to premature convergence and failure to maintain good population diversity. This paper proposes a Hénon map based adaptive PSO (HAPSO) for wavelet shrinkage image denoising. While significantly improving the population diversity of the particles, it also increases the convergence rate and thereby the precision of the denoising technique. The proposed PSO uses adaptive cognitive and social components and adaptive inertia weight factors. The Hénon map sequence is applied to the control parameters instead of random variables, which introduces ergodicity and stochastic property in the PSO. This results in a more improved global convergence as compared to the traditional PSO and classical thresholding techniques. Simulation results and comparisons with the standard approaches show the effectiveness of the proposed algorithm.

Index Terms - HAPSO, Hénon Map, Image Denoising, Premature Convergence, Wavelet Thresholding.

1. INTRODUCTION

Digital images have become indispensable, and with it the process of removing noise from an image while retaining its significant features has become an important requirement for analyzing images, especially in the field of digital image processing.

Ever since the work of Donoho [1]-[3] on wavelet thresholding, researchers have been trying to come up with techniques to make the thresholding process more adaptive. SureShrink, a wavelet Shrinkage technique is one such technique. Particle Swarm Optimization (PSO), a stochastic population based metaheuristic inspired from swarm intelligence, can be used to optimize the entire process of wavelet shrinkage image denoising. However, conventional PSO may get stuck in the local optima due to its fast convergence rate and even decrease the population diversity. To avoid this premature convergence and improve the population diversity, a Hénon map based adaptive PSO (HAPSO) approach is proposed. Here all the deciding parameters of PSO, the inertia weight, the cognitive

and social components, and the control parameters are made adaptive by special means. The use of Hénon map chaotic sequences for control parameters in PSO helps in escaping from the local minima. Hence it introduces ergodicity, irregularity and stochastic property in the PSO to improve the global convergence. The Adaptive PSO (HAPSO) is then used for selecting the optimum values for the parameters: wavelet threshold, type of wavelet basis and the level of decomposition; to denoise the digital image. In this paper, after providing the necessary background theory for classical wavelet shrinkage denoising technique and standard PSO, we give a detailed description of the proposed HAPSO technique.

The principal objective is to compare the performance of HAPSO, with standard PSO based wavelet shrinkage denoising technique and classical wavelet shrinkage denoising techniques for effective image restoration.

2. LITERATURE SURVEY

In recent years much advancement has been made in further optimizing the process of wavelet thresholding image denoising by using population based metaheuristics like Particle Swarm Optimization. Recently, a PSO-based approach was proposed by G. G. Bhatuda et al. [5] for learning the parameters of sub-band adaptive thresholding function for image denoising. In a different approach, modified Chaotic Particle Swarm Optimization was proposed by Xuejie Wang et al. [7] which used a chaotic PSO approach to optimize the wavelet threshold values. While to explore a complete solution space for suitable threshold, PSOShrink was proposed by Chan-Cheng Liu et al. [20]. Another variant of PSO, a Fast Particle Swarm Optimization was proposed by Du-Jin Liu et al. [9], for obtaining the most optimum wavelet threshold values. The earlier approaches, though generating high visual quality of images, have been more focused on optimizing a single wavelet parameter, the threshold values. In this paper, the HAPSO algorithm is proposed which uses an adaptive, ergodic version of PSO to optimize the wavelet threshold values, optimally select which wavelet basis to use from a set of mother wavelets and also, which level of decomposition to perform. More promising results are shown in terms of both PSNR and visual image quality as all the otherwise random parameters of PSO have been made adaptive in the proposed method.

3. WAVELET THRESHOLDING AND DENOISING

Wavelet thresholding is a nonlinear technique wherein an image or the given data is decomposed into wavelet coefficients. These detailed coefficients are then compared with a given threshold value, coefficients smaller than the threshold are set to zero while the others are retained or modified depending on the threshold rule. The image is then

¹Electronics and Communication Engineering Department, Indira Gandhi Institute of Technology, GGSIPU, Delhi, India
Email: ¹shruti4gandhi@gmail.com,

²doomra.sonal@gmail.com and ³akashtayal@yahoo.com

reconstructed from the modified coefficients, which is called *Inverse Discrete Wavelet Transform (IDWT)* [4]. Wavelet shrinkage denoising involves the following steps:

1. Acquire a noisy digital signal.
2. Compute a linear forward discrete wavelet transform of the noisy signal.
3. Perform a non-linear thresholding operation on the wavelet coefficients of the noisy signal.
4. Compute the linear inverse wavelet transform of the threshold wavelet coefficients.

Image denoising can be achieved by the following threshold techniques: VisuShrink[2] and SureShrink[3].

VisuShrink: Also known as Universal Shrinkage technique, this method follows the hard thresholding rule. The threshold value th is given as:

$$th = \sigma * \sqrt{2 * \log n} \quad (1)$$

where, σ^2 is the noise variance present in the signal and n represents the signal size or number of samples. The recovered images from VisuShrink are overly smooth and too many coefficients are removed as a result.

SureShrink: SureShrink, follows the soft thresholding rule and is smoothness adaptive, i.e., a threshold level is assigned to each dyadic resolution level by the principle of minimizing the Stein's Unbiased Risk Estimator for threshold estimates [3].

The threshold value for SureShrink is given by:

$$th^* = \min(th, \sigma * \sqrt{2 * \log n}) \quad (2)$$

where th^* denotes the value that minimizes Stein's Unbiased Risk Estimator, σ is the noise variance computed from Equation, and n is the size of the image.

4.0 STANDARD PARTICLE SWARM OPTIMIZATION

PSO is a metaheuristic, which imitates the social behaviour of natural organisms such as bird flocks and school of fishes to find a place with enough food.

The basic model consists of a swarm of N particles flying around in a D -dimensions search space, where each particle i is a candidate solution to the problem, and is represented by the vector $swarm(i)$ in the decision space. A particle has its own position and velocity, which represents the flying direction and step of the particle. The success of some particles will influence the behavior of their neighbors. The position of each particle $swarm(i)$ successively adjusts towards the global optimum depending on the two primary factors: The particle's best position $pbest_i$, denoted as $pbest_i = (p_{i1}, p_{i2}, \dots, p_{iD})$ and the whole swarm's best position, denoted as $gbest = (p_{g1}, p_{g2}, \dots, p_{gD})$. The difference between the current position of the particle i and the best position of its neighbourhood is represented by the vector $(gbest_i - swarm)$.

The following operations are applied to every particle for each iteration.

4.1 Updating velocity: Velocity defines the amount of change that a particle will undergo, it is given as:

$$v_{step}(t + 1) = w * v_{step}(t) + c_1 * r_1 * (pbest_i - swarm(t)) + c_2 * r_2 * (gbest_i - swarm(t)) \quad (3)$$

where w is the inertia weight, r_1 and r_2 are two random variables in the range $[0, 1]$ and the constants c_1 and c_2 represent the cognitive and social components, respectively. They represent the attraction that a particle has either toward its own success or toward the success of its neighbors. The direction and the distance followed by the particle are defined by the velocity.

4.2 Updating position: Each particle at every iteration updates its coordinates in the decision space according to the given equation.

$$swarm = swarm + v_{step} \quad (4)$$

Then it moves to the new position.

4.3 Updating the best found particles: Each particle updates (potentially) the best local solution:

$$If\ swarm(i) < pbest_i, then\ pbest_i = swarm(i) \quad (5)$$

Even the best global solution of the swarm is updated:

$$If\ swarm(i) < gbest_i, then\ gbest_i = swarm(i) \quad (6)$$

Therefore, with every new iteration, each particle changes its position according to its own experience and that of its neighboring particles [17] [19].

5.0 HÉNON MAP BASED ADAPTIVE PARTICLE SWARM OPTIMIZATION (HAPSO)

Standard PSO fails to come out of the local optimum of the solution space because of its fast convergence rate, resulting in poor global exploration. To improve the efficiency of searching and to balance the exploration and exploitation abilities of PSO, three important improvements over standard PSO have been introduced in the proposed algorithm.

5.1 Hénon map based Particle Swarm Optimization

The control parameters r_1 and r_2 are important parameters affecting convergence of PSO. The proposed algorithm uses Hénon equations to chaotically map these control parameters thereby introducing ergodicity, irregularity and stochastic property in PSO to get efficient results. Micheal Hénon introduced this map as a simplified version of the Poincare section of the Lorenz map [8].

The Hénon equations are given by:

$$x(t + 1) = 1 + y(t) - a * (x(t))^2 \quad (7)$$

$$y(t + 1) = b * x(t) \quad (8)$$

Here, the values of a and b are set to 1.4 and 0.3, respectively, because at these values the Hénon map has a strange attractor. The values of y are normalized in the range $[0, 1]$ and are assigned to the control parameters in the PSO equations, $r_1 = r_2 = y(t+1)$.

5.2 Adaptive Inertia Weight Factor (AIWF)

Inertia weight factor serves as the memory of the previous flight direction, preventing the particles from drastically changing their trajectory. The concept of linearly decreasing inertia weight factor over generations has long been introduced

and modified by Shi and Eberhart [15]. However, this paper uses adaptively varying inertia weight factor to achieve an optimum tradeoff between global exploration and local exploitation. The AIWF, $w(i)$, is determined as follows:

$$w(i) = \begin{cases} w(\min) + \frac{(w(\max) - w(\min)) * (f - f(\min))}{(f(\text{avg}) - f(\min))}, & f \ll f(\text{avg}) \\ w(\max), & f > f(\text{avg}) \end{cases} \quad (9)$$

Here, $w(\min)$ and $w(\max)$ are the minimum and maximum of w respectively, f is the current objective value of the particle, and $f(\text{avg})$ and $f(\min)$ are the average and minimum objective value of all the particles, respectively [6].

5.3 Adaptive Cognitive and Social Components

The cognitive component c_1 models the tendency of the particles to return to the previously found best solutions, whereas, c_2 , the social component measures the performance of the particles relative to its neighbors. It is known that larger value of c_1 as compared to c_2 leads to better searching of extremes in the entire search space at the early stage of the algorithm. While a larger c_2 as compared to c_1 ensures that the particles converge quickly to the global optimum value, later.

$$c_1 = c_{1Start} - \frac{\text{iter} * (c_{1Start} - c_{1End})}{\text{max_iter}} \quad (10)$$

$$c_2 = c_{2Start} + \frac{\text{iter} * (c_{2End} - c_{2Start})}{\text{max_iter}} \quad (11)$$

Here, c_{1Start} , c_{2Start} are the initial values of the learning factors and c_{1End} and c_{2End} are the final values. $Iter$ is the current iteration, and max_iter is the maximum number of iterations [10]. The above three approaches are then combined to give the Henon map based Adaptive Particle Swarm Optimization, described as follows:

$$v_{step}(t + 1) = w(i) * v_{step}(t) + c_1 * y(t + 1) * (pbest_t - swarm(t)) + c_2 * y(t + 1) * (gbest_t - swarm(t)) \quad (12)$$

where, $w(i)$ is the adaptive inertia weight factor (AIWF), $y(t+1)$ is the value of the normalized Hénon map sequence and c_1 and c_2 are the adaptive cognitive and social components.

Now this variant of PSO formed by combining all the above approaches, called the HAPSO is used for wavelet shrinkage image denoising.

6. WAVELET THRESHOLDING USING HAPSO

We use wavelet coefficient thresholding method for image denoising. For appropriate thresholding of coefficients several parameters have to be decided upon. They include:

- Choice of Wavelet
- Level of decomposition
- Threshold value at each level

We perform denoising of an image corrupted with Additive White Gaussian Noise (AWGN). HAPSO algorithm is employed to determine the optimum value of above mentioned parameters for denoising of noise affected images.

Parameters to be optimized		Permitted Values
Wavelet Basis	(Mother Wavelet)	Daubechies (db4, db6, db8), Coiflet (coif2, coif4)
Level of Decomposition		01 to 03
Threshold value		Range is from 0.5σ to 0.9σ ; σ is the estimate of noise variance. (up to three threshold values, one for each decomposition level)

Table 1: Allowed Values of Optimization Variables

Decomposition levels upto three is found to be most suitable for image denoising. PSNR is used as an evaluation criterion for measuring the effectiveness of the proposed technique.

The expression for PSNR is given as:

$$PSNR = 10 * \log_{10} \frac{G^2}{MSE} \quad (13)$$

where, G is the number of gray scale levels in the image and MSE is the mean square error between the estimated and the original image.

7. EXPERIMENTAL PROCEDURE

The main steps of HAPSO are described as follows:

Step1: Load a digital image and add additive white Gaussian noise (AWGN).

Step2: Give initial values to the parameters: w_{min} , w_{max} , $iter$, max_iter , c_{1Start} , c_{1End} , c_{2Start} , c_{2End} .

Step3: Initialize the velocities and position of each particle in the 5 dimensional search space using rand function (3 threshold values; one for each level of decomposition, a wavelet basis, a level of decomposition).

Step4: Apply Hénon map sequences (7) and (8) and assign the result to the control parameters r_1 and r_2 . This helps PSO to search for solution in wider space and jump out of the local minima.

Step5: Update the inertia weight factor according to equation (9). The adaptively varying inertia weight factor ensures global exploration in the beginning and local exploitation towards the end.

Step6: Update the cognitive and social coefficients according to the equations (10) and (11). This further enhances the convergence rate.

Step7: Update velocities and positions of each particle using equations (12) and (4), respectively.

Step8: Provide the required values to the wavelet shrinkage function to obtain the denoised image. Calculate the fitness value (PSNR) from the objective function. The PSNR of the denoised image is calculated using equation (13).

Step9: Compare the current fitness value of the particle with its previous best value (pbest). Update the value according to the equation (5).

Step10: Update the global best position in the swarm according to the equation (10).

Step11: Repeat steps 5 through 10 until maximum number of iterations reached.

The corresponding flowchart is shown in figure 1.

8. SIMULATION RESULTS

The HAPSO wavelet thresholding is applied on several natural grayscale images of size 256 X 256 and the simulations are performed in MATLAB 7.6.0 environment [11] [12]. The wavelet transform employs the most optimum wavelet basis, chosen from a set of mother wavelets (Daubechie 4, Daubechie 6, Daubechie 8, Coiflet 2, Coiflet 4) using HAPSO, at three levels of decomposition, the threshold value at each level of decomposition is also optimally decided by the proposed HAPSO. In this paper the particle's search space is spread over five dimensions. To assess the performance of HAPSO, it is compared with the standard PSO, VisuShrink and SureShrink for two different test images. The adaptive parameters c_1 , c_2 , $w(i)$, r_1 and r_2 require their respective variables to be initialized. These variables are initialized with the values that have been found to have the most optimum effect on the proposed method.

- i. $a = 1.4, b = 3$; used in equations (7) and (8)
- ii. $w(min) = 0.9, w(max) = 0.3$; used in equation (9)
- iii. $c_{1Start} = c_{2End} = 2$; used in equations (10) and (11)
- iv. $c_{2Start} = c_{1End} = 0.3$; used in equations (10) and (11)

The maximum number of iterations for all the images is kept at 50 and the swarm size of the particles also at 50. Here, PSNR is used to evaluate the performance of the proposed algorithm, represented as the cost function. Higher the value of the PSNR, the better is the effectiveness of the method used for image denoising. Equation (13) is used for evaluating the PSNR. The two cost functions graphs for 'cameraman.tif' are shown in Figure 2 and Figure 3. Figure 2 is for image denoising using standard PSO, whereas Figure 3 is for image denoising using the proposed HAPSO. On comparison, the proposed algorithm HAPSO clearly proves to be more effective than standard PSO.

The Cost function/ Fitness function becomes constant as it reaches the value 71.8 for PSO based wavelet shrinkage denoising while it reaches 74.184 for HAPSO based wavelet shrinkage denoising for 50 iterations.

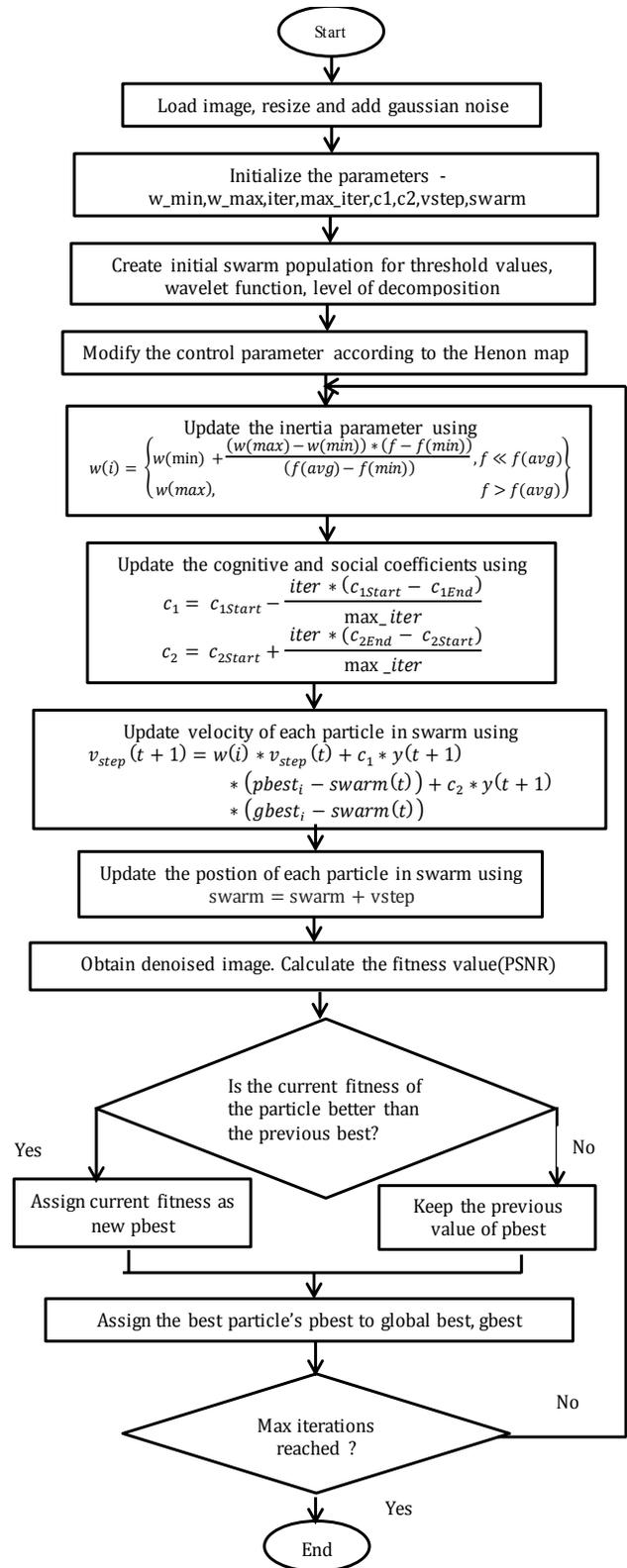


Figure 1: HAPSO Flowchart

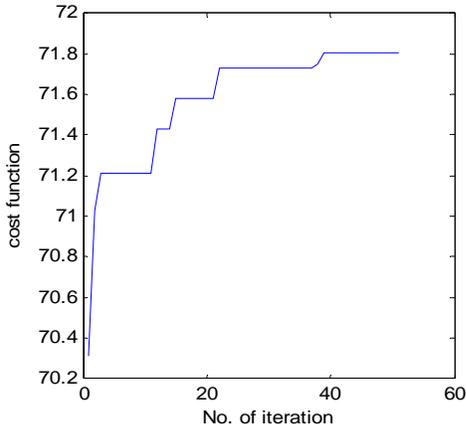


Figure 2: Cost Function for standard PSO

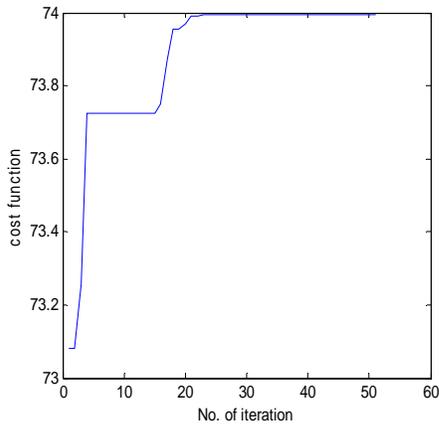


Figure 3: Cost function for HAPSO

9. CONCLUSION AND FUTURE SCOPE

In this paper, a Hénon map based adaptive Particle Swarm Optimization method for wavelet shrinkage thresholding is presented. Here we have shown that the use of HAPSO for wavelet shrinkage image denoising is more efficient when compared to the standard PSO based wavelet shrinkage denoising, and far exceeds the classical methods VisuShrink and SureShrink in terms of, not only the PSNR value but also the visual quality. Here, the HAPSO was used to optimize three wavelet parameters, the threshold value, the wavelet basis and the level of decomposition. The elements of

adaptability and ergodicity introduced through Hénon equations to chaotically map the control parameters. Followed by the use of adaptive inertia weight factor (AIWF) and adaptive cognitive and social coefficients, this variant of PSO makes the process even more effective by greatly improving the global exploration of solution space, population diversity and rate of convergence. As a result, we get more optimized values of the parameters required for denoising the image using wavelet shrinkage. A tabular comparison of the simulation results and PSNR values of the denoised images as shown in Table 2, proves the effectiveness of the proposed HAPSO. The work can be further extended by using modified adaptive hybrid metaheuristics combining Genetic Algorithm, PSO, Simulated annealing or their different combinations to obtain a more optimal set of wavelet thresholding parameters, and to further improve the convergence behavior as well as the quality of the denoised image.

REFERENCES

- [1] D. Donoho, "De-Noising By Soft Thresholding", IEEE Transactions on Information Theory, Vol. 41, pp. 613-627, 1995.
- [2] David L. Donoho and Iain M. Johnstone, "Ideal Spatial Adaptation by Wavelet Shrinkage," Biometrika, Vol. 81, No. 3, Aug., 1994, pp. 425-455.
- [3] David L. Donoho and Iain M. Johnstone, "Adapting to Unknown Smoothness via Wavelet Shrinkage," Journal of American Statistical Association, 90(432):12001224, December 1995.
- [4] Mallat S.G, "A Theory for Multiresolution Signal Decomposition: the Wavelet Representation," IEEE Trans. Pattn Anal. Mach. Intell., 11, 674- 693, 1989.
- [5] Bhutada, G. G., R. S. Anand, and S. C. Saxena. "PSO-based learning of sub-band adaptive thresholding function for image denoising." *Signal, Image and Video Processing* 6.1 (2012): 1-7.
- [6] Liu, Bo, Ling Wang, Yi-Hui Jin, Fang Tang, and De-Xian Huang. "Improved Particle Swarm Optimization Combined With Chaos." *Chaos, Solitons & Fractals* 25, no. 5 (2005): 1261-1271.
- [7] Wang, Xuejie, Yi Liu, and Yanjun Li. "A Novel Wavelet Threshold Optimization via PSO for Image Denoising." In *Computational Intelligence and Design (ISCID), 2011 Fourth International Symposium on*, Vol. 1, pp. 352-355. IEEE, 2011.
- [8] Coelho, Leandro dos Santos, and Viviana Cocco Mariani. "A Novel Chaotic Particle Swarm Optimization Approach using Hénon Map and Implicit Filtering Local Search for Economic Load Dispatch." *Chaos, Solitons & Fractals* 39, no. 2 (2009): 510-518.
- [9] Du-jin Liu, Siming Li, Shuxia Sun, Zhaoyu Ding. "Application of Fast Particle Swarm Optimization Algorithm in Image Denoising". *Recent Advances in Computer Science and Information Engineering*

Shrinkage Methodology Used	PSNR cameraman.tif	PSNR bird.gif
VisuShrink	20.6876	22.7382
SureShrink	23.7427	26.2087
PSO based Wavelet shrinkage	71.8	72.028
HAPSO based Wavelet shrinkage	74.184	77.812

Table 2: Final Comparison of PSNR values

- lecture Notes in Electrical engineering, Volume 126, 2012, pp. 559-566.
- [10] Yongshou Dai, Hui Niu, “An Improved PSO Algorithm and its Application in Seismic Wavelet Extraction,” I.J. Intelligent Systems and Applications, 2011, Vol. 5, pp. 34-40.
- [11] MATLAB 7.6.0, Image Processing Toolbox.
- [12] MATLAB 7.6.0, Wavelet Toolbox.
- [13] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, “Image Coding using Wavelet Transform,” IEEE Trans. Image Processing, vol. 1, no. 2, pp. 205– 220, 1992.
- [14] Rafael C. Gonzalez and Richard E. Woods – Digital Image Processing; Second Edition; Pierson Prentice Hall, 2007.
- [15] Yuhui Shi and Russell Eberhart, Indiana University Purdue University Indianapolis, “A Modified Particle Swarm Optimizer”, 1998.
- [16] Magnus Erik Hvass Pedersen, Hvass Laboratories, Technical Report no. HL1001, “Good Parameters for Particle Swarm Optimization,” 2010.
- [17] El-Ghazali Talbi , “Metaheuristics from Design to Implementation”, University of Lille – CNRS – INRIA
- [18] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, “Image Coding Using Wavelet Transform,” IEEE Trans. Image Processing, vol. 1, no. 2, pp. 205– 220, 1992.
- [19] Kennedy, J. and Eberhart, R., “Particle Swarm Optimization,” Proceedings of the IEEE International Conference on Neural Networks, Perth, Australia 1995, pp. 1942-1945.
- [20] Chan-Cheng Liu, Tsung-Ying Sun, Shang-Jeng Tsai, Yu-Hsiang Yu, Sheng-Ta Hsieh, Heuristic wavelet shrinkage for denoising, Applied Soft Computing, Volume 11, Issue 1, January 2011, Pages 256-264

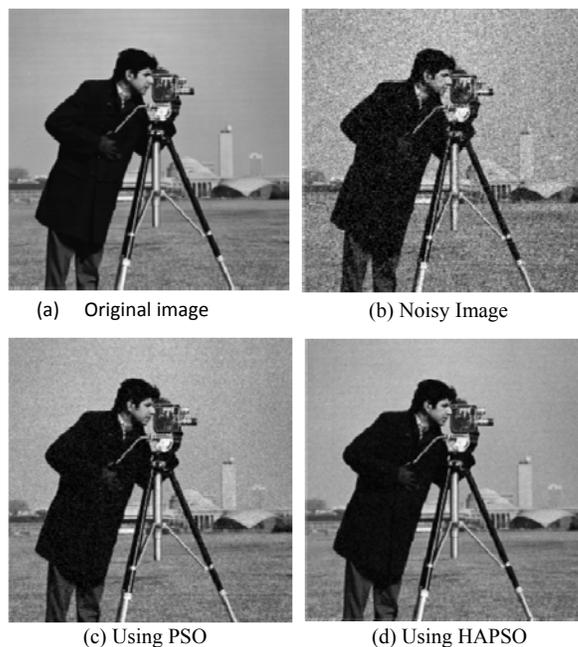


Figure 4: Simulations performed on ‘cameraman.tif’

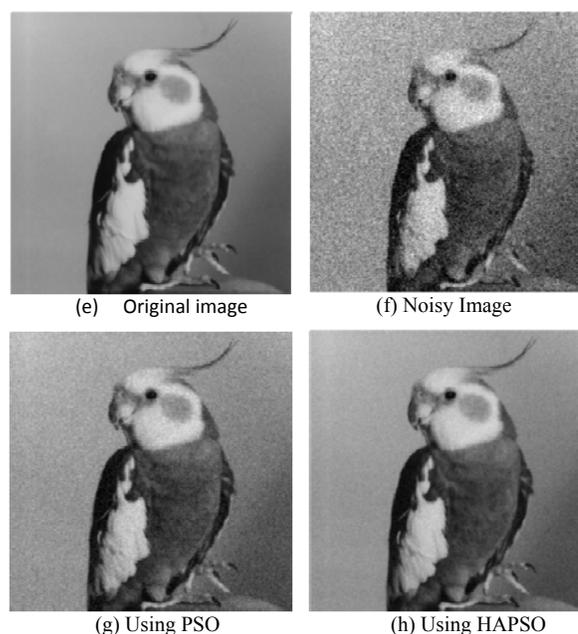


Figure 5: Simulations performed on ‘bird.gif’

BIJIT - BVICAM's International Journal of Information Technology
(A Half Yearly Publication; ISSN 0973 - 5658)

Subscription Rates (Revised w.e.f. January, 2012)

Category	1 Year		3 Years	
	India	Abroad	India	Abroad
Companies	Rs. 1000	US \$ 45	Rs. 2500	US \$ 120
Institution	Rs. 800	US \$ 40	Rs. 1600	US \$ 100
Individuals	Rs. 600	US \$ 30	Rs. 1200	US \$ 075
Students	Rs. 250	US \$ 25	Rs. 750	US \$ 050
Single Copy	Rs. 500	US \$ 25	-	-

Subscription Order Form

Please find attached herewith Demand Draft No. _____ dated _____
For Rs. _____ drawn on _____ Bank
in favor of **Director, "Bharati Vidyapeeth's Institute of Computer Applications and
Management (BVICAM), New Delhi"** for a period of 01 Year / 03 Years

Subscription Details

Name and Designation _____

Organization _____

Mailing Address _____

_____ PIN/ZIP _____

Phone (with STD/ISD Code) _____ FAX _____

E-Mail (in Capital Letters) _____

Date:

Signature

Place:

(with official seal)

*Filled in Subscription Order Form along with the required Demand Draft should be sent to the
following address:-*

Prof. M. N. Hoda

Editor-in- Chief, BIJIT

Director, Bharati Vidyapeeth's

Institute of Computer Applications & Management (BVICAM)

A-4, Paschim Vihar, Rohtak Road, New Delhi-110063 (INDIA).

Tel.: 91 - 11 - 25275055 Fax: 91 - 11 - 25255056 E-Mail: bijit@bvicam.ac.in

Visit us at: www.bvicam.ac.in/bijit

INDIACom-2014

8th INDIACom; 2014 International Conference on

Computing for Sustainable Global Development

(05th-07th March, 2014)

IEEE Conference Record Number # 32558

INDIACom-2014 is aimed to invite original research papers in the field of, primarily, Computer Science and Information Technology and, generally, all interdisciplinary streams of Engineering Sciences, having central focus on sustainable computing applications, which may be of some use in enhancing the quality of life and contribute effectively to realize the nations' vision of sustainable inclusive development using Computing. INDIACom-2014 is an amalgamation of four different international conferences which will be organized parallel to each other, as parallel tracks. These are listed below:-

Track #1: International Conference on Sustainable Computing (ICSC-2014)

Track #2: International Conference on High Performance Computing (ICHPC-2014)

Track #3: International Conference on High Speed Networking & Information Security (ICHNIS-2014)

Track #4: International Conference on Software Engineering & Emerging Technologies (ICSEET-2014)

INDIACom-2014 will be held at **Bharati Vidyapeeth, New Delhi (INDIA)**. The conference will provide a platform for technical exchanges within the research community and will encompass regular paper presentation sessions, invited talks, key note addresses, panel discussions and poster exhibitions. In addition, the participants will be treated to a series of cultural activities, receptions and networking to establish new connections and foster everlasting friendship among fellow counterparts.

Full length original and unpublished research papers based on theoretical or experimental contributions related to the following topics, but not limited to, are solicited for presentation and publication in the conference:-

- Algorithms and Computational Mathematics
- Green Technologies and Energy Efficient Systems
- IT for Education, Health & Development
- IT for Environmental Sustainability
- IT for Sustainable Agriculture Development
- IT for Water Resources Management
- IT for Consumers' Right
- IT for Crisis Prevention & Recovery
- IT for Disaster Management and Remote Sensing
- IT for other day to day problems
- E-Governance
- Knowledge Management
- E-Commerce, ERP, CRM & Knowledge Mining
- Technology for Convergence
- Distributed and Cloud Computing
- Parallel, Multi-core and Grid Computing
- Reconfigurable Architectures
- Changing Software Architectural Paradigms
- Programming Practices & Coding Standards
- Software Inspection, Verification & Validation
- Software Sizing and Estimation Techniques
- Agile Technologies
- Artificial Intelligence and Neural Networks
- Computer Vision, Graphics, and Image Processing
- Modelling and Simulation
- Embedded Systems and Robotics
- Human Computer Interaction
- Databases
- Data Mining and Business Intelligence
- Big Data Analytics
- Operating Systems
- Data Communication, Computer Networks and Information Security
- Wireless Networking
- Network Monitoring Tools
- Next Generation Internet
- Mobile Computing
- Entertainment Technologies
- Multimedia Computing
- Information and Collaboration Systems
- Fuzzy and Soft Computing
- Bioinformatics
- Medical Informatics
- Education Informatics
- Computational Finance
- Research Methods for Computing
- Case Studies & Applications

Paper Submission

Authors from across different parts of the world are invited to submit their papers. Authors wishing to submit their papers must refer to the website, for paper structuring and formatting guidelines in detail, at <http://www.bvicam.ac.in/indiacom/Technical%20Guidelines.asp>. Authors should submit their papers online at <http://www.bvicam.ac.in/indiacom/loginReqSubmitPaper.asp>. Unregistered authors should first create an account on <http://www.bvicam.ac.in/indiacom/addMember.asp> to log on and submit paper. Only electronic submissions will be considered. E-Mailic submissions will not be considered.

Review Process, Publication and Indexing

The conference aims at carrying out two rounds of review process. In the first round, the papers submitted by the authors will be assessed on the basis of their technical suitability, scope of work and plagiarism. The corresponding authors of qualifying submissions will be intimated for their papers to be double blind reviewed by at-least two experts on the basis of originality, novelty, clarity, completeness, relevance, significance and research contribution. The shortlisted papers will be accepted for presentation and publication in the conference proceedings, having **ISSN 0973-7529** and **ISBN 978-93-80544-10-6** serials. Conference proceedings will also be available in the form of CD-ROMs. All accepted papers, which will be presented in the conference, will be submitted for inclusion to **IEEE Xplore**, as a part of **IEEE's Conference Publication Programme**, subject to their terms and conditions. Further details are available at www.bvicam.ac.in/indiacom.



**Bharati Vidyapeeth's
Institute of Computer Applications
& Management (BVICAM)**

A-4, Paschim Vihar, Rohtak Road, New Delhi-63 (INDIA)

Technically Sponsored by



Supported by



CSI Region-I &
CSI Division-I



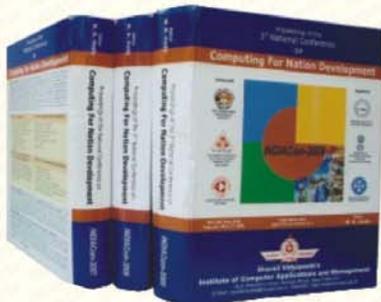
GURU GOBIND SINGH
INDRAPRASTHA UNIVERSITY



IE (I), Delhi Centre



ISTE, Delhi Section



(Copies of the proceedings of past INDIAComs)

Correspondence

All correspondences related to the conference must be sent to the address:-

Prof. M. N. Hoda

General Chair, INDIACom - 2014

Director, BVICAM, A-4, Paschim Vihar, New Delhi -63 (INDIA)

Tel.: 91-11-25275055, TeleFax:91-11-25255056, 09212022066 (Mobile)

E-Mails: conference@bvicam.ac.in, indiacom2014@gmail.com

visit us at: <http://www.bvicam.ac.in/indiacom>